NBER WORKING PAPER SERIES

MOBILITY FOR ALL:
REPRESENTATIVE INTERGENERATIONAL MOBILITY ESTIMATES OVER THE 20TH CENTURY

Elisa Jácome
Ilyana Kuziemko
Suresh Naidu

Working Paper 29289
http://www.nber.org/papers/w29289

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2021

We thank Ahna Pearson and Paola Villa-Paro for excellent research assistance. Sandy Black, Leah Boustan, Ellora Derenoncourt, James Feigenbaum, Nathan Hendren, Tom Hertz, Chi Hyun Kim, Trevon Logan, Chris Muller, and Zachary Ward have provided invaluable data and feedback at various stages of this project. We thank seminar participants at ASSA, Berkeley, Cologne, Georgetown, NBER, Northwestern, NYU Wagner, Princeton, Toronto, UCL, USC, Warwick, and Wharton. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w29289.ack

Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century
Elisa Jácome, Ilyana Kuziemko, and Suresh Naidu
NBER Working Paper No. 29289
September 2021
JEL No. H0,J15,J16,N3

ABSTRACT

We present the first estimates of long-run trends in intergenerational relative mobility for samples that are representative of the full U.S.-born population. Harmonizing all surveys that ask about father's occupation and own family income, we develop a mobility measure that allows for the inclusion of non-whites and women for the 1910s–1970s birth cohorts. We show a robust increase in mobility between the 1910s and 1940s cohorts, about half of which is driven by absolute convergence in racial income gaps. We also find that excluding Black Americans, particularly Black women, considerably overstates mobility throughout the 20th century.

Elisa Jácome
Stanford Institute for Economic Policy Research
Stanford University
John A. and Cynthia Fry Gunn Building
366 Galvez Street
Stanford, CA 94305
ejacome@stanford.edu

Ilyana Kuziemko
Department of Economics
Princeton University
239 J.R. Rabinowitz Building
Princeton, NJ 08544
and NBER
kuziemko@princeton.edu

Suresh Naidu
Columbia University
420 West 118th Street
New York, NY 10027
and NBER
sn2430@columbia.edu

1 Introduction

Intergenerational relative mobility—how tied an individual's place in the income distribution is to her parents' place in the income distribution while she was growing up—has long been an object of interest, especially in the United States. A high level of mobility is viewed as an important part of Americans' identity as a nation: that it is a "land of opportunity" even for those who grew up poor. American perceptions of intergenerational mobility remain significantly higher than measured mobility and correlate with decreased support for redistribution (Alesina *et al.*, 2018).[1] In political philosophy, intergenerational relative mobility is widely viewed as a proxy for equality of opportunity (Roemer and Trannoy, 2015) and the overall fairness of a society.

The rise in inequality since the 1970s and 1980s has further increased interest in intergenerational mobility (Mogstad and Torsvik, 2021). Cross-country variation suggests that in modern data high levels of inequality correlate with greater income persistence between parents and children (Corak, 2013), raising the concern that the rise in income inequality over the past generation will lead to a decline in intergenerational mobility in the US. However, direct evidence on the correlation between inequality and mobility across *time* within the US has been more limited. As Song *et al.* (2020) write: "evidence of long-term trends in intergenerational mobility is largely absent" (p. 251).[2]

The main contribution of this paper is simple: we present, to the best of our knowledge, the first estimates of long-run intergenerational relative mobility trends for a *representative* sample of U.S.-born individuals. In particular, we show mobility estimates for children born in the 1910s through the 1970s.[3] As Table 1 shows, a handful of papers have made important contributions to our understanding of long-run trends in intergenerational relative mobility. However, for data reasons, they include only subsets (and typically *advantaged* subsets) of the population. Song *et al.* (2020) shows mobility of occupational prestige from 1830 to 1980, but only for white men. Using a clever synthetic-panel strategy based on the status information conveyed by first names, Olivetti and Paserman (2015) can compare occupational mobility between fathers and sons to that of fathers and sons-in-law, but only for white men and married white women. Collins and Wanamaker (2017) and Ward (2020) include Black Ameri-

[1] The idea that upward mobility could reduce support for redistribution among a majority of *rational* voters was explored theoretically in Benabou and Ok (2001), though they emphasize that the result only holds for voters with very low levels of risk aversion.

[2] Similarly, Mazumder (2018, p. 225-226) write: "One active topic of research that has not yet been resolved is whether there have been major changes in intergenerational mobility in the United States over time."

[3] Note that we do not examine intergenerational *absolute* mobility, which captures the probability that a child's income as an adult surpasses her parents' income (in real dollars) while she was a child. For recent work on intergenerational absolute mobility, see Chetty *et al.* (2017) and Berman (2018).

cans, but only men. Papers on intergenerational relative mobility that include women and non-whites present results only for more modern periods (see, e.g., Chetty *et al.*, 2014b, Solon, 1992, Chetty *et al.*, 2020, or Mazumder, 2018) or short snapshots of time (see, e.g., Card *et al.*, 2018).

The goal of including women and non-whites motivates our methodology. Instead of relating occupational status of one generation to the next (which complicates looking at women, as few worked in the historical period), we introduce a family-income-to-income-score mobility concept. Essentially, we relate the self-reported family income (adjusted for inflation) of the adult child to an *income score* meant to predict her family income while she was growing up, based on characteristics such as her race, her father's occupation, and her region. Family income is readily reported in many datasets, and is a question that male as well as female respondents can answer. Moreover, it will naturally pick up income gaps by race.

We locate (to the best of our knowledge) all surveys that ask individuals their current family income and their father's occupation while they were growing up, ranging from well-known surveys like the General Social Survey to more obscure ones like "Americans View Their Mental Health" and "The National Survey of Black Americans." All of these surveys also ask race and the region (at the very least, South versus elsewhere) where the respondent was born or grew up. Our baseline estimates use the 1940 Census to predict family income for fathers using *occupation* $\times$ *race* $\times$ *South* cells (though we show robustness to many variations in predicted family income, including using the father's education whenever available). We then relate self-reported family income of the adult child to her predicted income while growing up, estimating both intergenerational elasticity and rank-rank mobility relationships. Pooling and harmonizing surveys decreases the sampling noise in each cohort, resulting in more precise and stable estimates than would be found in any single survey.

Our main finding is that both IGE and rank-rank correlations fell (meaning that mobility rose) between the 1910s and 1940s birth cohorts. The IGE (rank-rank correlation) falls from 0.58 (0.39) for those born in the 1910s to 0.36 (0.24) for those born in the 1940s. Between the 1940s and 1970s birth cohorts, the IGE measure drifts upward again, while there is little change in the rank-rank correlation. While survey-based measures of mobility are less reliable for these later (1960s–1970s) cohorts, the u-shape we find for the IGE tracks measures of inequality in the family income of the adult children in our sample, suggesting a cross-cohort "Great Gatsby Curve" relating mobility estimates to inequality, as has been found in cross-country data (Corak, 2013). On the other hand, the absence of the same increase for the rank-rank measure also offers evidence against the tight linkage of inequality and mobility. Of course, just as with the cross-country correlation, we make no claims to causality between the adult children's inequality measures and their mobility.

As noted, our paper is motivated by the absence of historical mobility measures that are representative of the full U.S.-born population. In the second part of the paper, we focus on subgroups (mostly the four subgroups defined by Black/white race and male/female sex) and in particular how movements of these subgroups contributed to (or slowed) the increase in mobility from the 1910s to 1940s cohorts. Like inequality measures and unlike means, the full-population IGE (or rank-rank) slope is *not* a weighted average of subgroup slopes. In particular, the between-subgroup differences in parental *mean* income play a major role. We make this point more precise in the paper, but the basic logic is shown in Figure 1. In this case, the full-population IGE is greater than the IGE measure for *either* subgroup, because group B comes from such an extremely disadvantaged part of the parental-income distribution and remains disadvantaged in adulthood.

In our subgroup analysis, we show several key changes in the mapping of parental income to own adult income among our four subgroups and how they relate to the overall change in full-population mobility. From the 1910s–1920s to the 1940s–1950s birth cohorts, there is substantial catch-up for Black Americans, with their entire regression line shifting upward by roughly twenty-five log points (or seven rank percentiles). Whites also enjoy income growth in real terms, and their IGE and rank-rank slopes become flatter (meaning that, *within the white population*, parental income matters less in predicting own adult family income). Our decomposition, applying Hertz (2008) to our historical data, shows that the Black-white catch-up in levels of income accounts for roughly half of the rise in mobility (and the flattening of the white slope accounts for the remainder). This result is quite striking given that Black Americans are just over ten percent of the U.S. population. But given that they are drawn from an extreme part of the parental income distribution (in their case, an extremely low part) and in our historical period did not intermarry with whites, changes in their adult outcomes will be statistically influential on estimates of the full-population regression line.

In this paper we pay particular attention to Black women. First, because of data limitations, there has been almost no work on historical intergenerational mobility that includes this group. Second, as we will demonstrate later in the paper, because Black women tend to grow up in the bottom of the income distribution (as do their male counterparts) and in our historical period are the lowest-income group as adults (even poorer than Black men), they play an especially outsize role in increasing full-population intergenerational persistence measures. As just one example, in 1920, the rank-rank correlation increases from 0.27 to 0.32 (with non-overlapping confidence intervals) when Black women are added to the rest of the sample. Excluding even this small share of the population overstates early twentieth-century U.S. mobility considerably.

In the final part of the paper, we compare the mobility patterns by race and sex in our historical period to those documented in modern data (individuals currently in their 30s) by Chetty *et al.* (2020). The modern data show significant retrenchment in the progress made by Black Americans (as does our data for the 1960s and 1970s cohorts). One reading of this reversion to earlier trends is that the Black-white convergence of the mid-century cohorts (who entered the labor market during and immediately after the Civil Rights era) was merely a temporary deviation from steady-state gaps. However, the patterns by gender suggest a more dynamic story. Over most of our sample period, Black women are the poorest group, with much lower family income than their male counterparts. But by the end of our sample period (the 1970s birth cohort) this gap has closed and in the modern data it has in fact flipped, with family income for Black women being substantially higher than that for Black men in Chetty *et al.* (2020). Thus, a careful examination of mobility by race *and gender* over the 20th century suggests important shifts over time, which future work might link to policy or other changes.

The remainder of the paper is organized as follows. In the next section we describe the various datasets we use. In Section 3, we describe our methodology, in particular the family-income-to-income-score mobility concept. In this section we describe in detail how we calculate income scores to approximate parental income. Section 4 presents our results for the full, representative population. Section 5 presents a decomposition of the mobility measures and decomposes the rise in mobility into differential mobility by race and gender. Section 6 relates our findings by race and sex to modern data. Section 7 concludes.

2 Data

In this section, we briefly describe the datasets that we use in this paper and share summary statistics. Far greater detail can be found in Appendix D.

2.1 Datasets and sampling rules

We have located to the best of our knowledge all surveys that ask respondents their current family income, their fathers' occupation while they were growing up (with sufficient detail), their race, and the region of the country where the respondent was born or grew up (at least to the level of South versus other regions). We end up locating 15 different surveys, and details on all of them are provided in Appendix D. Most readers will be familiar with some (e.g., the General Social Survey or the American

National Election Survey), but others are not as well known (e.g., the National Survey of Black Americans or Americans View their Mental Health).

In some cases, the data we use are in fact panel datasets that follow individuals and families over time (e.g., the Panel Study of Income Dynamics [PSID] and the National Longitudinal Surveys of Mature Women and Older Men) and have often been used to measure mobility for more modern periods. To remain consistent within our methodology, however, we do not use the *panel* components of these datasets. In the first wave, these panel datasets often ask the adult respondent questions about their own childhood, and it is this linkage that we use to estimate the respondent's family income in childhood.

Following the IGE literature, we restrict attention to U.S.-born men and women in the 30–50 age range in order to ensure that we are measuring life-cycle earnings as closely as possible. While some recent papers have not limited themselves to ages close to forty, in all cases we limit ourselves to this age range. Papers that use occupation scores or education levels of the adult child may well have less worry about life-cycle bias as these measures may be more stable across adulthood, but because we are using self-reported family income, we take care not to stray too far from prime-age years.

In many cases, the data collection for these surveys was explicitly meant to be representative and provides survey weights to correct deviations due to sampling error. In those cases, we use the provided sampling weights. Of course, some of these surveys target one sex (e.g., the National Fertility Survey) or one race (e.g., National Survey of Black Americans) and so are clearly not representative of the full U.S.-born population. In the early cohorts, we also have a substantially lower share of women in our data relative to the general population. For this reason, we will always re-weight the pooled dataset so that each cohort has weighted shares for white women, white men, Black women and Black men of 0.44, 0.44, 0.06 and 0.06, respectively.[4] In the appendix, we show that our main results barely change under other weighting schemes, including not weighting at all.

2.2 Summary statistics

The first panel of Table 2 shows summary statistics of the fathers of the respondents in our main dataset, separately by decade of birth. In this table we do not weight at all, so that readers can get a sense of the raw data.

[4] We only focus on individuals whose race is classified as white or Black. Individuals of other races account for tiny shares of the surveys' samples in these historical time periods (1% or less of the sample in the pre-1950 cohorts). The decomposition in Section 5 also highlights that groups with very small population shares are unlikely to affect the full-population measures of persistence.

The decline of agriculture as a dominant occupation for fathers is readily apparent for children in the 1910s–1950s birth cohorts, falling from over one-third to less than one-tenth. As noted, we do not have father's education in every survey, but the table shares summary statistics from those surveys that do include father's education. In our earliest birth cohorts, the fathers in our data are born in the last few decades of the nineteenth century and thus grew up before the high school movement, which is reflected in their low levels of secondary education. Less than twenty percent of the fathers of our 1910s and 1920s birth cohorts graduated from high school. College graduation was a rarity for these fathers and as late as the 1950s birth cohorts less than one in six of respondents have fathers who completed college.

Summary statistics for the adult children (i.e., the survey respondents) appear in the second panel of the table. The age of respondents is relatively similar and always close to forty, as we would expect from our 30–50 age restriction. In contrast to past historical work which focused on whites (white men, in fact), our samples have coverage of Black individuals very close to their population shares. Past work that has applied linkage techniques to the Black population in the Census have also tended to result in samples somewhat smaller than the population shares (Collins and Wanamaker, 2017, Ward, 2020).

A number of trends among the children in our data merit comment. The rise of educational attainment from the 1910s to the 1950s birth cohorts is striking. High school attainment increases from one-half to 90 percent, and college graduation rates nearly triple from ten to twenty-eight percent. The increase in education from one generation to the next is massive as well: for the 1910s to 1930s birth cohorts, the likelihood our survey respondents graduate from high school is triple that of their fathers.

Another marked trend for the adult children in our data is the decline in veteran status (which the table reports only for men in surveys that asked about veteran status). While over seventy percent of men in our 1920s cohort report military service, by the 1950s cohort military service has become quite rare. Finally, another noticeable trend is union membership: while it holds steady in the high-twenties to low-thirties for our early cohorts, it begins a steady decline with the 1950s cohort.

Table 3 separates our data (unweighted, as in the previous table) by time period, race and sex and compares it to the relevant population in the Census. As before, we see that in all periods and separately for men and women, our data are very close to representative on race (roughly ten to fifteen percent of the sample). In fact, one of the only variables on which there are small discrepancies between our raw survey data and the Census data is education in the earliest birth cohorts. For example, whites in our data have a high-school completion rate about ten percentage points higher than their Census counterparts (the differences are positive but slightly smaller for Black individuals). This difference is smaller for all groups in later birth cohorts.

Otherwise, our raw survey data is remarkably similar to the Census in terms of age, the share living or originating from the South (an especially important variable for Black respondents), and marital patterns. This table emphasizes the fact that, practically speaking, it is simply easier to gather representative data in a single cross-section (as our surveys do) than to maintain a representative sample over many years via panel data (whether the connections over time are created by Census linking or explicitly following the same person longitudinally as in the PSID or National Longitudinal Surveys [NLS]).

3 Methodology

In this section, we introduce a new mobility concept that we can calculate for all U.S.-born Americans from twentieth-century birth cohorts. We assess the ability of father's *income score* to predict the adult child's *family income*. Our income score predicts father's household income using occupation, race, and Southern region (at minimum) and in some cases education as well. So, importantly, we do not assign the same income scores to white and Black respondents. Our use of the adult child's family income instead of the child's occupation in adulthood allows us to incorporate women as well, given that many women did not work in the historical period but can still report a measure of family income to survey enumerators.

3.1 Calculating father's income scores

IPUMS provides 1950-based occupational income scores that go back as far as 1850, which calculate the median total income of the people (pooling men and women) in each occupation in 1950. We modify the standard IPUMS *occscore* variable in a number of ways.

First, not all of our surveys give father's occupation categories that are as detailed as those in the Census. Across all of our surveys, we can harmonize occupations into 28 categories (corresponding to the categories in the ANES). We thus build and use crosswalks that map Census occupations into these 28 categories. These coarsened bins include broad occupations like doctors, clerical workers, craftsmen, and farm laborers, and are listed in Appendix D.

Second, when constructing income scores, we limit the Census samples to men between the ages of 30 and 50 who are living with a biological child younger than 18 years old. This sample restriction should better proxy household income of *fathers* with a given occupation, which is the population of interest when we try to predict income during the respondent's childhood.

Third, we use the 1940 instead of 1950 Census (though we show robustness to using the 1950-based occupational income scores, as well as robustness to many other modifications of the family income score). The 1940 Census better captures the pre-Great-Compression wage and income distribution, a point made by Collins and Wanamaker (2017) and Ward (2020). Moreover, the full-count sample of the 1940 Census is available, whereas only the one-percent sample of the 1950 Census is currently public and only sample-line respondents are asked about income. The larger sample size of the 1940 Census helps reduce noise when estimating the median income of smaller cells.

Finally, unlike the standard Census *occscore* variable, we take the median household income of this sample of fathers by occupation, race (Black versus white) and region (South versus elsewhere). We follow recent papers that attempt to include Black Americans, such as Collins and Wanamaker (2017) and Ward (2020), who make similar adjustments. Given widespread discrimination and occupational segregation, using occupational scores computed from pooled Black and white populations likely mismeasures occupational incomes. In order to focus on Black-white differences in relative mobility, we thus generate occupation scores separately by race. Moreover, the South is far poorer than other regions during our sample period, so pooling across all regions throws out valuable information, especially for Black respondents who are vastly overrepresented in the region.

One feature of historical measurement of occupational incomes is that farm income is notoriously difficult to impute, as it is both highly volatile (being subject to weather and price shocks) as well as difficult to measure (as comprehensive measurement of agricultural costs is difficult to capture). Moreover, the 1940 Census income variable excludes income from self-employment, which includes most farmers. We therefore follow the approach of Collins and Wanamaker (2017) to calculate the income of farmers in 1940, using the income of farm laborers in 1940 as well as the ratio of farmer-to-farm-laborer income in the 1960 Census to impute the income of 1940 farmers. We similarly adjust the income of self-employed businessmen in 1940 using a similar approach. More detail on these adjustments is available in Appendix D.

Figure 2 compares the standard 1950 Census *occscore* variable (on the y-axis) to our income score (on the x-axis), all in 1950 dollars. Not surprising, income scores of Black individuals are almost always to the left of whites, and in particular Black Southern income scores (gray diamonds) are to the left of Black income scores from other regions (pink circles). A Southern income-score penalty exists for whites as well, with the blue solid triangles (white Southerners) typically to the left of the green solid circles (white non-Southerners).

Appendix Figure A.1 shows the Gini coefficient based on our predicted parental income measure, separately by respondents' birth decades. We cannot directly compare these Gini coefficients to those from, say, the 1940 or 1950 Census, as our measure will pick up none of the inequality coming from within-cell variation. Nonetheless, the measure does capture the known decline in inequality from the 1930s to 1950s.[5]

While we show robustness to several modifications of this income score later in the paper, this measure serves as our baseline childhood household income score, as we can calculate it for the respondents in all fifteen of our surveys. We briefly foreshadow some of the adjustments we make to predicted childhood income here, and defer details to Section 4. A natural question is the validity of our 1940-based occupation scores for earlier cohorts, especially for farmers (Feigenbaum, 2018). In addition to checking the robustness of our results to dropping respondents with fathers in agriculture, we also use earlier data sources to measure farm income (namely, the 1900 Census of Agriculture to construct race-by-region farmer income) as well as non-farm income (namely, the 1901 Cost of Living Survey). For surveys in which father's educational attainment is available, we show robustness to augmenting our *occupation* × *race* × *South* scores to *occupation* × *race* × *South* × *education* scores to improve our prediction of parental income during childhood.

3.2 Comparison to past measures of parental income

Data limitations have long plagued the study of mobility in the United States, and our approach is no exception. We briefly review the main approaches in the literature, highlighting their advantages and disadvantages to better put our approach and results in context.

Papers using historical data

An advantage of studying older cohorts is that the Census provides de-identified data for those in the 1940 and earlier Censuses (and will de-identify the 1950 Census next year). Recent papers have used linking algorithms to find the same individual across Censuses based on their name, year of birth and place of birth. Ferrie (1996) was an early and important contribution to this literature.

However, this approach is not without complications and limitations. First, there is an active literature on the correct linking methodology and the preferred tolerance for

[5] In this graph, inequality is high and declining slightly from the 1910s to the 1930s. Given limited income data from this period, it is difficult to compare our measure to any "ground truth" from the Census or other sources.

rates of falsely matching and missing true matches (see, e.g., Abramitzky *et al.*, 2019 and Bailey *et al.*, 2020). Matching methodologies are still in flux and best practices will likely evolve as machine-learning techniques improve. Second, at least with the current available technologies, the linked population is not representative of the full population. Most obviously, the linked sample is not representative by sex, as women often change their names upon marriage and thus a representative group of women cannot generally be linked. To date, all published mobility papers using Census linking drop all women. Even beyond gender, certain types of names are very hard to link with precision. For example, there are too many John Smiths born in New York State in any given year to know with confidence that men with those characteristics in two different Census years are in fact the same person; conversely, long, foreign names are often changed, preventing matches.

An important example of a group which proves challenging for Census linking is Black Americans. For example, an important contribution of Ward (2020) is the inclusion of Black men, but his linked sample is only two-percent Black before those observations are up-weighted in the subsequent analysis. Similarly, Collins and Wanamaker (2017) are able to find reliable adult matches for three and five percent of Black children in the 1880 and 1900 Census, respectively. Moreover, Black Americans, and particularly Black men, are systematically under-counted in Censuses even before any linking is performed.[6] In short, linking historical Censuses proves quite difficult beyond white men.

As income is not available until 1940, most mobility work using Census data focuses on the occupational status of the father (as we do, though we adjust it along additional dimensions). Relative to a single snapshot of parental or father's income, which is a very noisy proxy for average childhood income and thus leads to severe attenuation bias (Solon, 1992), a single snapshot of father's occupation may have the advantage of being more stable over time. But a single observation of a father's occupation has noise from two sources, as Ward (2020) recently highlights. First, fathers change occupations from year to year, especially when occupations are measured at the three-digit level that is often used in this literature. While this attenuation bias is likely smaller than that from year-to-year changes in family income, it could still be substantial. Ward (2020) shows that estimates that measure mobility using father's occupation as observed in a single Census substantially over-estimate mobility relative to those that use multiple observations across different Censuses. Second, Census-takers appear to record occupation with substantial error, at least in the historical period. As Ward

[6] O'Hare (2019) calculates that the net under-count rate for the Black population has gone from 8.4% in 1940 to 2.5% in 2010.

(2020) details, in a special case when a re-census was required in St. Louis in 1880, one-third of occupations were reported differently only five months later, despite the reference date for the occupation being unchanged.

Given the challenges of linking, researchers have turned to creative solutions. As already noted, Olivetti and Paserman (2015) use a synthetic panel of first names, which allows them to examine (married) women as well as men (though they only include whites). To the extent that children stay in their parents' households as adults, then household surveys like the Census allow researchers to observe both child and parents *without* needing to link, an insight Card *et al.* (2018) and Hilger (2015) have used to study intergenerational mobility with respect to education. But this approach only works for periods in which most children have completed their education while living with their parents and of course does not provide a workable solution when the outcome of interest is the adult child's family income, as few children remain with their parents during their prime-age years.

Our approach in many ways circumvents the challenges associated with linking. As our data come from simple cross-sections, they tend to be representative, as it is much easier for a survey to find a representative cross-section in a given year than it is to maintain representativeness following a given sample across time or identifying linkages across historical Censuses. Indeed, as we already discussed, the percent of Black respondents in our (unweighted) data is very close to that in the full U.S. population, even for our earliest cohorts.

That said, there are important subgroups that may be missed by surveys. Given our focus on representativeness of the U.S. population, especially by race, the fact that incarcerated or otherwise institutionalized people are unlikely to complete the surveys in our sample may bias our estimates of intergenerational mobility. Appendix Figure A.2 shows the share of individuals ages 30–50 who are institutionalized (e.g., in correctional facilities or mental hospitals), separately by subgroup and cohort. The stark increase in the Black male incarceration rate for cohorts born since the 1960s is clear in the Census data. But there is little differential trend for Black male institutionalization for those born prior to 1960, which are the cohorts that are the focus of our study.

We do not observe fathers for a single year, but rather observe them in the recollections of their adult children during their prime-age years. In that sense, we do not face the problem that Census researchers face of having the bad luck of observing the father in a particularly unrepresentative year in terms of his occupation. It seems natural to assume that the adult child would remember the occupation her father mostly did, so the retrospective nature of our data likely aids in identifying the main occupation of the father.

Papers using more modern data

For those interested in studying more modern cohorts, two data sources are especially useful. First, some datasets have been collected with the express purpose of measuring intergenerational mobility, such as the PSID and the NLS datasets. Second, IRS data allow linking of a small number of cohorts (those born around 1980).

The PSID and NLS datasets have many advantages for modeling intergenerational mobility. First, they tend to have multiple observations (five or even ten) of father or family income while the child is growing up, alleviating concerns about attenuation bias. They have been fielded over decades, so the children can now be observed in their prime-age years. Because many advantaged children spend much more time in formal education, their earnings tend to be disproportionately depressed in the late twenties and early thirties relative to their prime-age earnings, so measuring the adult child's income at these ages also leads to downward bias of persistence measures. Haider and Solon (2006) offer as a rule-of-thumb to observe children as close to age forty as possible.

However, it is difficult for long panels such as these to avoid attrition, which typically results in non-representative samples as the most disadvantaged respondents prove harder to track over time and across generations. Schoeni and Wiemers (2015) show that the patterns of attrition by parent and child income results in biased estimates of intergenerational mobility. Indeed, as we show in Appendix Table A.1, individuals for whom we observe five or even ten years of childhood household income in the PSID have fathers who are slightly whiter and much more educated than the general population of fathers. Notwithstanding these concerns, recent work has found a decline in mobility between the 1950s and 1960s birth cohorts, consistent with our results (see results from Davis and Mazumder, 2017, using NLS datasets).

Chetty *et al.* (2014b) pioneered the use of administrative data, available since the 1990s, to study U.S. mobility. These data obviate the need for linking (the observations have identification numbers), and are much less susceptible to attrition or life-cycle bias, as many years of income of both parents and children are available. Even with these administrative data, roughly seven percent of children cannot be linked to parents for various reasons, cohorts born earlier than the late 1970s cannot be examined, and even for the cohorts considered, parental income is only observed while the individuals are adolescents.[7]

Relative to IRS data, our sample sizes are orders of magnitude smaller, preventing us from breaking the data into neighborhoods or single percentiles as in Chetty *et al.*

[7] See Heckman *et al.*, 2013, Uguccioni, 2021, as well as cites therein for evidence that *early* childhood resources are especially important to later-life outcomes.

(2014a), Chetty and Hendren (2018a) and Chetty and Hendren (2018b). And while our approach allows us to reach further back in history than IRS or PSID data, it cannot reach as far back as Census linking (as in Ward, 2020, Song *et al.*, 2020, Collins and Wanamaker, 2017 or Olivetti and Paserman, 2015). The types of surveys we use only became common in the 1940s and 1950s and thus cohorts born before 1910 cannot be studied without violating our sampling rule of including only individuals between ages 30 and 50. And of course, IRS and PSID data have the important advantage of directly observing parental income (the IRS with the added advantage of third-party verification in many cases), whereas we rely on adult children's recall, an issue we now discuss in greater detail.

3.3 Assessing adult children's recall of parents' occupation

Perhaps the key challenge for our analysis is that it depends on the accuracy of the adult children's responses when asked to report their fathers' occupations. In Appendix C, we perform a number of exercises that we hope will bolster readers' confidence in the accuracy of our income score, while acknowledging we cannot fully validate the measure.

First, we show that the occupations reported by male and female respondents tend to match (as we would expect, given that there is no documented evidence of sex selection in the US in our historical period and thus boys and girls on average grow up in the same families). Specifically, we compare the estimated income scores—which are of course based in part on recall of father's occupation—for men versus women for the full sample (by decade) as well as for white and Black respondents, finding essentially no significant differences beyond what would be expected by chance and in all cases the differences by sex are small and switch signs (see Appendix Tables C.1 and C.2).

Second, we show that the average income score as well as the mix of occupations reported by our survey respondents are similar to the occupations of actual fathers in the Census when these respondents were growing up (see Appendix Tables C.4–C.6). This exercise helps alleviate concerns that children tend to inflate the status of their father when they are asked to recall their upbringing.

Finally, we use the PSID to perform a direct validation. Again, more details are provided in Appendix C, but we provide a short recap here. We make use of the fact that (beginning in 1997) the PSID asks household heads to recall their father's occupation while they were growing up and in many cases we directly observe the fathers of these respondents in earlier waves of the surveys (i.e., the 1960s and 1970s) when they are asked to report their own occupations. Over 80 percent of these household heads report an occupation that the father also reports and the most common mistakes are small and understandable (e.g., one party reporting "craftsmen" and the other reporting "operatives").[8] Indeed, the correlation between logged income scores based on father's self-reported occupation and those based on the child's recall is 0.85 and the relationship is very linear across the entire support of father's income scores (so we do not see, for example, that the children of the lowest-status dads tend to overstate their father's occupational status or that children of the highest-status dads tend to understate). Further, the coefficient from a regression of the five-year average of log father's income on our constructed income score (from the retrospective question), results in a coefficient very close to 1, suggesting that our retrospective income scores are quite closely correlated with father's actual permanent income. While we cannot directly verify that the accuracy of adult children observed in the 1990s and early 2000s in the PSID is a good proxy for the accuracy of our survey respondents in earlier years, it is reassuring that in a setting where we can directly validate adult children's recall, it appears highly reliable.

3.4 Specifications

We estimate variants of the following two specifications, both standard in the mobility literature. We begin with the classic log-log specification estimated in Becker and Tomes (1979):

$$\log(y_{ic}) = \beta^c \log(y^p_{ic}) + f(\text{age}_i) + \delta_y + \epsilon_i. \tag{1}$$

In this equation, β^c is an estimate of the IGE for cohort c. We control for a quadratic in the age at which we observe the adult child (though recall we already restrict the sample to be observed at ages 30–50, which should limit life-cycle effects). We also include survey-year fixed effects in all specifications.

Next, we follow Chetty *et al.* (2014a) and calculate ranks for fathers and children. The rank of the father is the percentile (based on the income score described in the previous subsection) among all fathers having a child in cohort c. Similarly, the rank of the child is the rank of family income among all children born in cohort c. The mapping of child's rank to parent's rank (the copula of the joint distribution) tends to be linear and can handle zeros, which may be missed in the (logarithmic) IGE specification. Chetty *et al.* (2014a) focus on this specification:

$$\text{Rank}_{ic} = \gamma^c \text{Rank}^p_{ic} + \delta_y + \epsilon_i. \tag{2}$$

In this estimation, γ^c is an estimate of the rank-rank correlation for cohort c.

[8] Note that we often see multiple observations of father's self-reported occupation, as household heads were asked about their current occupation during each survey wave.

In the final panel of Table 2, we show these income and rank measures, where incomes are all in 1950 dollars. There is only minor top- and bottom-coding of the adult children's family income in each birth decade. Real family income of the children grows robustly over the 1910s–1940s birth cohorts, consistent with strong post-war economic growth. Fathers' income *scores* grow more slowly, as by construction they can only represent occupational upgrading across time (because they are based on the 1940 income distribution, though we revisit this assumption in robustness checks).

An obvious but important point is that with sufficient granularity in the income measures, the average rank for fathers and children must both be fifty. Reassuringly, we see an average close to fifty across all of our birth decades. Missing information for parental income is a challenge even for modern, administrative data, and so we include the share of children who have missing information for father's income, which rises especially after 1950. We do not include these observations in most of the analysis, but later we show robustness to adding them back into the sample with various assumptions about missing fathers' incomes or by using mothers' occupations when available.

4 Results for representative samples

The main finding we describe in this section is a rise in intergenerational relative mobility between the 1910s and 1940s cohorts. In the next section we try to understand this trend by splitting up the sample by gender and race, but in this section we merely aim to establish the robustness of this main result.

4.1 Main results

The first series of Figure 3 shows the IGE for survey respondents over time, pooling across surveys and applying our baseline population-adjusted weights. We show the IGE separately by decade of birth and report the corresponding estimates in Table 4. Between the 1910s and 1940s birth cohorts, the IGE falls markedly, from roughly 0.58 to 0.36. We then see an increase in this measure in subsequent birth cohorts, so that the IGE appears to take on a u-shape over time.

The second series shows the results from the rank-rank specification, which mirrors very closely those of the IGE in terms of trends. As is typically found in other papers, our rank-rank coefficients are lower in magnitude than our IGEs: it begins the sample period just under 0.40 and declines to a low of just under 0.24 for the 1940s birth cohorts. Like the IGE, it also rises after 1940, though in a significantly less pronounced manner.[9]

For several reasons, we focus on the *decline* in the IGE and rank-rank measures that occurs from the 1910s to the mid-century birth cohorts, instead of the subsequent rise in the IGE thereafter. First, as we noted in our discussion of Table 2, the share of data with missing information about fathers increases over time, so trends toward the latter part of our sample period might reflect sample selection.[10] Second, beginning with the 1960s birth cohorts, modern panel data such as the PSID and later on linked administrative IRS data become available, so we feel our relative contribution to understanding mobility patterns in the modern period is smaller.

Figure 4 shows the decline in intergenerational persistence between the 1910–1919 cohorts and the 1940–1949 birth cohorts as bin-scatters figures. The first panel shows the change in the IGE relationship: a large shift upward (reflecting real income growth relative to parents across the parental income distribution) as well as a significant flattening of the slope (because the upward shift is especially large among individuals growing up with less family income).

The second panel of Figure 4 shows that the decline in the rank-rank is also large and precisely estimated. Given that by construction there can never be an overall increase in rank (its average must always be 50) we see only a flattening of the slope. While caution is warranted in terms of comparing the levels of our rank-rank estimates (which use parental income scores) to those from modern administrative data (which use actual income data averaged over several years from the parents), we use the modern estimates as rough benchmarks. The rank-rank slope we find for the 1910s–1920s cohorts is roughly equal to the modern US (see Chetty *et al.*, 2014a), whereas the slopes we find for mid-century cohorts are close to the modern estimates in Canada and Denmark (see Connolly *et al.*, 2019 and Helsø, 2021, respectively).

Finally, we plot a third series in Figure 3, the Gini coefficient based on the self-reported family income for the respondents in our surveys (that is, the adult children in the parent-child pair).[11] Quite strikingly, the inequality and persistence measures move in tandem over these birth cohorts. As we noted in the Introduction, much of

---

9 Appendix Figure A.3 plots the estimates separately for each survey in order to give readers a sense of which surveys contribute to each decade's estimate and their relative magnitudes. We do note, however, that when we separate surveys in this manner, the individual estimates are less likely to be representative, as they might only reflect individuals of one race, one sex, or younger/older age groups. For example, men born in the 1950s in the National Longitudinal Survey of Young Men are between the ages of 30–31, which likely explains their relatively lower estimates (i.e., they are more affected by life-cycle bias).

10 However, to the extent that the estimates in this later period suffer from measurement error, then if anything, the u-shape over this entire time period will be even more marked.

11 Appendix Figure A.4 plots the Gini coefficient of family income using the decennial Censuses, confirming that the cohorts born in the 1930s and 1940s experienced the least inequality in their own adult family income.

the support for the hypothesis that inequality and mobility are inversely related (the so-called Great Gatsby relationship) comes from cross-country data. Our results in Figure 3 provide some evidence for the Great Gatsby curve using variation over time within the US. To date, such evidence has been lacking.

Interestingly, Song *et al.* (2020) find a similar effect to our results for cohorts born at mid-century, though they include only white men in their analysis. They note that while their long-run results suggest general stability among birth cohorts in the twentieth century, a potential exception is the "drop in the intergenerational correlation estimated from pooled social surveys for the 1950 cohort (born between 1946 and 1955), namely the early baby-boom generation." While they do not emphasize it as much as the 1946–1955 decline, their persistence correlations are also lower for the 1936–1945 and the 1956–1965 cohorts, which we also find.[12] Song *et al.* (2020) write that "we consider the deviation of the 1950 birth cohort best interpreted as suggestive. Proper interpretation of this deviation awaits future research with further evidence." We view our results as adding yet another piece of evidence in support of a temporary but significant increase in mobility for mid-century U.S. birth cohorts. Similarly, including white and Black men, Ward (2020) finds that mobility is significantly lower in 1920 than in 1950 (though he does not have data for the intervening years) consistent with our results for representative samples.[13]

In summary, we have so far provided evidence of a significant decline in IGE and rank-rank persistence measures between the 1910s and 1940s birth cohorts. Importantly, these results reflect samples that are representative of the full U.S.-born population, including women and non-whites. In the subsequent subsections, we attempt to show robustness of this result to what we consider the most central concerns.

4.2 Corroborating evidence from respondent's education

On average, more educated individuals have significantly higher earnings and family income, with the exact return to education varying over time. Thus, it would be

[12] As Song *et al.* (2020) note, other papers finds hints of such a result as well. Using data from the GSS, Hout (1988) find rising intergenerational mobility of occupational status from the early 1970s to the mid-1980s, which would correspond to some of our most mobile birth cohorts.

[13] A natural question that might arise is how our estimates compare to those in the literature. Appendix Figure A.5 shows how the estimates vary as we transition from using the IPUMS *occscore* variable for both generations among white men (which is closer to the traditional estimates in the literature), to incorporating income scores that vary by race and region for each generation, and finally to using reported family income. We see that the decline in IGE estimates is present, albeit less marked, when using income scores for both generations. However, the decline become more salient once we allow there to be within-occupation variation in incomes for the adult children that is not captured by racial or regional differences.

somewhat surprising if the predictive power of parental income on children's education did not fall given that its predictive power over children's family income did.

To examine this idea, we estimate variants of equations (1) and (2) where in both cases we put the adult child's self-reported years of schooling as the outcome variable (available in all of our datasets). Figure 5 shows the results from both of these estimations, as usual, by birth decade. The relationship between father's income and respondent educational attainment declines sharply between the 1910s and 1950s birth cohorts. Appendix Figure A.6 illustrates these changes using bin-scatter figures, highlighting that this weakening relationship is largely driven by the rapid increase in respondents' high school completion in the bottom half of the income distribution, rather than the later rise in college completion.

Using data from the modern period, Landersø and Heckman (2017) has questioned whether mobility is truly lower in places such as Scandinavia than in the US, because when education of the adult child is the outcome of interest instead of earnings or income, mobility measures in the US and Scandinavia look more similar. In our analysis, both family income and years of education appear to have a decreasing dependence on childhood income over the first half of the twentieth century.

4.3 Adjusting for father's education

As noted earlier, we proxy the child's family income while she was growing up by taking the median of 1940 Census household income for fathers by occupation, race and region (South versus elsewhere). Our hope is that this measure can proxy for the respondent's long-run income during childhood. The extent to which it does *not* pick up idiosyncratic, mean-zero variation in family income from year to year is in fact a strength, as it reduces attenuation bias. But if it misses systematic variation in family income not picked up by *occupation* $\times$ *race* $\times$ *South*, then it will lead to bias. Moreover, it is not *a priori* obvious the sign or trend over time of the bias.

As noted earlier, for more than half of our surveys, respondents were also asked about their fathers' education. We can thus augment the fathers' income scores by predicting income by *occupation* $\times$ *race* $\times$ *South* $\times$ *education* for respondents in these surveys. Father's education is one of the most important reasons why family income could systematically deviate from our *occupation* $\times$ *race* $\times$ *South*-based income score. Indeed, adding information about education significantly increases the power of our income scores to predict 1940 family income (the R-squared rises from 20.6 to 24.3).

The first panel of Appendix Figure B.1 compares the IGE with the original *occupation* $\times$ *race* $\times$ *South* income score—using the baseline sample as well as the sub-sample of respondents who are asked fathers' education—to the IGE using these augmented scores in this restricted sub-sample. The three series are very comparable in both levels and trends: in particular, they show the marked decline between the 1910s and 1940s birth cohorts. The second panel shows that the decline in the rank-rank measure is also unchanged by augmenting parental income score with father's education. Thus, when we significantly improve our income scores with an important predictor, the trends in mobility remain unchanged, providing some reassurance that systematic, unobserved within *occupation* $\times$ *race* $\times$ *South* cell variation in income is not driving our results.[14]

4.4 Adjustments to farm income

Past work (e.g., Collins and Wanamaker, 2017, Song *et al.*, 2020, Ward, 2020) on historical mobility has paid special attention to farmers, both because they are such a large part of the population in the nineteenth and early twentieth centuries and because their income is difficult to model. For example, it is well known that the relative position of farmers declined from the late nineteenth to mid-twentieth century. Using data from the 1950 U.S. Census and the 1915 Iowa Census, Feigenbaum (2018) finds that the 1950 IPUMS *occscore* is a good predictor of actual income in 1915, with the important exception of farmers. Farmers experienced a substantial fall in relative income between 1910 and 1950 (for example, in 1915 Iowa, farmers made median income, but in the full 1950 U.S. population, farmers were at the bottom of the income distribution).

As noted earlier, our baseline income scores acknowledge the difficulty in using income data from the 1940 Census to measure farm income, and thus follow the approach in Collins and Wanamaker (2017) to adjust farmers' incomes. In our baseline measure, for example, white respondents born in the 1910s–1920s cohorts outside of the South and who have farmer fathers are estimated to be growing up around the 25$^{\text{th}}$ percentile of the income distribution.

Nevertheless to test for the robustness of the main result to this baseline adjustment for farmer income, we begin by using an alternative source of data to calculate farmer income. Specifically, we follow the approach in Goldenweiser (1916) and Abramitzky *et al.* (2012) and use the 1900 Census of Agriculture to calculate farmers' net earnings. In our calculations, we allow for variation at the *race* $\times$ *South* level and take into account the share of each group that is not farm owners (i.e., part owners, or cash or share tenants). Using this data source tends to increase farmer fathers' rank in the income distribution for the earliest cohorts (e.g., by roughly 10 rank percentiles for

[14] Appendix Figure B.2 shows how the IGE and rank-rank estimates evolve as we transition from occupation-based income scores to income scores that also include information about a father's race, region, and education. If the decline were driven by an increase in measurement error across cohorts, then we would expect the decline to decrease in magnitude as we add important predictors to the income scores. Instead, we see that the decline is persistent and robust across various ways of predicting a father's income.

white farmers outside of the South). Appendix Figure B.3 shows that our main result of a marked decline in persistence remains when we use this alternative data source.

As an alternative approach, also shown in Appendix Figure B.3, we simply drop farmers to ensure that our mobility patterns are not being entirely driven by this population, a population for which it is hard to estimate childhood income. Again, the conclusion that mobility increased substantially between the 1910s and 1940s birth cohorts is unchanged.

4.5 Alternatives to household income in the 1940 Census

We also check the robustness of our main result to alternative ways of constructing income scores. We begin by using father's income—as opposed to household income—for fathers with a given occupation, race, and Southern residence. Appendix Figure B.4 shows that the IGE, and especially the rank-rank estimates, hardly change and the decline in persistence remains.

We then consider using an alternative data source, not only for farmers, but for all occupations. We combine our estimates from the 1900 Census of Agriculture with average earnings from the 1901 Cost of Living Survey and again adjust for variation at the *race* × *South* level (Preston and Haines, 1991). The third series in this figure shows that despite using completely different data sources—measured forty years apart in time—for both agricultural and non-agricultural occupations, the increase in mobility remains marked between the 1910s and 1940s cohorts. These alternative income scores also allow us to account for changes in the relative status of certain occupations over the first half of the twentieth century. For example, in the earliest cohorts, children whose fathers were semi-skilled operative and kindred workers have ranked childhood income around the 35$^{\text{th}}$ percentile of the income distribution with this alternative measure, compared to the 45–50$^{\text{th}}$ percentile using our baseline measure.

Next, we combine our data sources, so that fathers are assigned income scores using the data source closest in time to when the respondent is 10 years old. That is, the 1910s–1930s cohorts are assigned income scores that are weighted averages of the 1900- and 1940-based income scores, and all subsequent birth cohorts are similarly assigned income scores that are weighted averages of scores constructed using the 1950–1990 Censuses. Again, the general patterns for both the IGE and rank-rank estimates remain unchanged.[15]

[15] It is worth noting that our IGE and rank-rank estimates are always lowest in 1940, which is closest in time to the Census that we use to construct our baseline income scores. To the extent that our pre-1940 baseline estimates are attenuated down due to measurement error, then we are likely understating the decline between the 1910s and 1940s cohorts. Indeed, once we use either earlier sources or the most contemporaneous data sources possible, the decline in the IGE appears even more marked.

Finally, as discussed in Section 3.1, while we believe that our revisions to the IPUMS *occscore* methodology are valuable and appropriate, the final series in this figure shows robustness to merely using the *occscore* variable to predict childhood household income.

4.6 Other robustness checks

Table 2 shows that the information needed to calculate predicted childhood income is not available for all respondents. This situation arises almost always because the respondent does not report father's occupation (presumably because she doesn't remember, chooses not to report it, or grew up without her father). In Appendix Figure A.7 we incorporate the roughly 2,000 respondents whose fathers were present but not working (e.g., retired, unemployed), imputing their income scores using the ratio of household incomes of working and non-working fathers in the 1960 Census. The next series in this figure instead incorporates the roughly 4,700 respondents who provided information about their mother's occupation, assigning them income scores based on mothers who were household heads in the 1940 Census. The estimates hardly change, but the precision of the estimates for later cohorts tends to increase (given that the share of respondents with no information on the father but available information on the mother's occupation is largest in these last cohorts). We also show robustness to a particular extreme assumption about respondents who provided no information about either parent: that their household had zero income, or in other words, that their family had the lowest possible rank for predicted childhood income.

Appendix Figure A.8 shows the robustness of the main result to alternative weighting schemes: namely, using the provided survey weights without any additional adjustments for population shares and to using no weights at all. In all of these checks, we continue to find an increase in mobility between the 1910s and 1940s cohorts.

One concern with this long-term view of mobility might be that mortality rates were high for early cohorts (for men and Black men in particular; see e.g., Preston *et al.*, 2003), so that selection into the sample might be changing over time. If individuals who grow up in poorer households are those with higher mortality rates, then if anything, this decline in mortality would likely bias us against finding a decline in intergenerational persistence between the 1910s and 1940s birth cohorts (in the earliest cohorts, individuals born poor who are still alive at prime age would likely be positive selected, which would bias intergenerational persistence downward). Nevertheless, we still take seriously this consideration and Appendix Figure A.9 compares our baseline results to those that focus on individuals ages 30–45 and 30–40, both of which are less affected by differential mortality rates. The rise in mobility is unchanged in these sub-samples.

Finally, another notable demographic change that took place in the 20th century was the change in household size for the mid-century cohorts (i.e., the Baby Boom). Appendix Figure A.10 adjusts the family income of adult children using information about household size as well as the interpolated fathers' income scores using the median household size in that *occupation* × *race* × *South* cell when the respondent is 10 years old. We see that this adjustment does not affect the rise in mobility between the 1910s and 1940s birth cohorts.

5 Decomposing the rise in mobility

In this section, we show how to decompose the overall IGE and rank-rank relationships into factors related to subgroups, building on Hertz (2008). We then use this decomposition to show how much changes in mobility or income among subgroups, particularly by race and gender, explain the overall decline in IGE and rank-rank coefficients that we found in the previous section.

5.1 Decomposing the IGE and rank-rank slopes

Consider any partition of the full sample, emitting subgroups $g \in G$ with subgroup g's share of the total sample given by p_g. Further, let β_g^{IGE} be equal to β from estimating the IGE equation $y_i^c = \alpha + \beta y_i^p + e_i$ (where, as usual, y^c and y^p are the adult child's and the parent's log income, respectively) on the subgroup g.

From the OLS formula and the law of total covariance, the whole-population IGE is given by:

$$\begin{aligned}
\beta^{\text{IGE}} =& \frac{Cov(y^c, y^p)}{Var(y^p)} \\
=& \frac{1}{Var(y^p)}\Big(E_g[Cov(y^c, y^p)] + Cov(E[y^c \mid g], E[y^p \mid g])\Big) \\
=& \underbrace{\sum_{g \in G} p_g \frac{Var(y^p|g)}{Var(y^p)} \beta_g^{\text{IGE}}}_{\text{Weighted average of subgroup slopes}} + \underbrace{\frac{Cov(E[y^c|g], E[y^p|g])}{Var(y^p)}}_{\text{Between-group covariance of subgroup averages}},
\end{aligned} \tag{3}$$

where E_g denotes the expectation over groups g.

A slight modification gives a similar expression for the whole-population rank-rank slope. In this case, β_g^{RR} is the β from $r_i^c = \alpha + \beta r_i^p + e_i$ (where, as usual, r^c and r^p are the within-cohort rank of the adult child's and the parent's income) on the subgroup g. Assuming that both the parental and adult children's ranked incomes have a uniform distribution, the same application of the law of total covariance gives:[16]

$$\begin{aligned}\beta^{RR} =& \sum_{g\in G} p_g \frac{Var(r^p|g)}{Var(r^p)}\beta_g^{RR} + \frac{Cov(E[r^c|g], E[r^p|g])}{Var(r^p)} \\ =& 12 \times \Big(\sum_g p_g Var(r^p|g)\beta_g^{RR} + \sum_g p_g \mathbf{E}[r^p|g]E[r^c|g] - 0.25 \Big)\end{aligned} \tag{4}$$

To ease intuition and to focus on one of the key applications for our paper, we rewrite the IGE decomposition for two groups, W and B:

$$\begin{aligned}\beta^{\text{IGE}} =& p_W \frac{Var(y^p|W)}{Var(y^p)}\beta_W^{\text{IGE}} + (1-p_W)\frac{Var(y^p \mid B)}{Var(y^p)}\beta_B^{\text{IGE}} \\ &+ \frac{p_W \mathbf{E}[y^p \mid W]\cdot \mathbf{E}[y^c \mid W] + (1-p_W)\mathbf{E}[y^p \mid B]\cdot \mathbf{E}[y^c \mid B] - \mathbf{E}[y^p]\mathbf{E}[y^c]}{Var(y^p)}.\end{aligned} \tag{5}$$

The decomposition helps clarify two points. First, because population shares act as weights in the first two terms of equation 5, changes in the within-group IGE of the majority group will, perhaps not surprisingly, affect the full-population IGE. Second, and less obviously, the decomposition highlights the important role of between-group differences in parental income y^p in determining the full-population IGE. To see this point, assume for the moment that W and B are two distinct subgroups, but are drawn independently *from the same distribution* of parental income y^p. Thus, there exists no between-group variation in y^p and $Var(y^p) = Var(y^p \mid W) = Var(y^p \mid B)$ and $\mathbf{E}[y^p] = \mathbf{E}[y^p \mid W] = \mathbf{E}[y^p \mid B]$. In this special case of no between-group differences in parental income, $\beta^{\text{IGE}} = p_W\beta_W^{\text{IGE}} + (1-p_W)\beta_B^{\text{IGE}}$, or in other words, the full-population IGE is the average of the two subgroup IGE slopes weighted by the subgroup share of the total population. This result holds regardless of the adult childhood outcomes (e.g., even if the mean adult income y^c of group B is well below that of group W).[17] However, if there exist large differences in parental income between the two groups (as there is in our context, when the two groups are Black and white Americans), then this third term will be heavily weighted and will play a key role in determining the full-population IGE.

[16] If either r^c or r^p do not have a uniform income distribution, then one can simply use the decomposition in equation 3, substituting y^p with r^p, y^c with r^c, and β_g^{IGE} with β_g^{RR}.

[17] In the less-extreme case in which the two groups just have the same average parental income (i.e., $\mathbf{E}[y^p] = \mathbf{E}[y^p \mid W] = \mathbf{E}[y^p \mid B]$), then the third-term still cancels out, and the full-population IGE is a weighted average of the subgroup IGEs, where the weights are a function of population shares and (conditional and unconditional) variances of parental income.

5.2 Mobility by race and gender

As we noted in the Introduction and in the decomposition shown above, an important reason to examine the mobility of representative samples is that relating mobility measures of subgroups to the full-cohort mobility is complicated. In this subsection, we show the mappings of father to adult children's incomes separately by the race and gender of the respondent and how these mappings change over time. We then use the decomposition in equation (3) to show how changes in relative mobility and in the income of these groups contributed to (or stymied) the overall rise in mobility that we find from the 1910s to the 1940s birth cohorts.

Given the discussion above, we expect that the between-group component will prove important for a decomposition along racial subgroups, given that Black Americans grow up with far less parental income than whites in our period (as well as today). By the same logic, however, any decomposition by gender will be very different. Men and women grow up, on average, in the same households in the US (recall that we confirm this claim empirically in Appendix Table C.1 and show it also holds separately for white and Black respondents in Appendix Table C.2). Thus, the between-group component of equation (5) should be close to zero and the full-population IGE is well approximated by the simple mean of the within-gender IGE slopes (as men and women each represent roughly half of the population). Put differently, the male-only IGE will be a biased measure of full-population IGE only if the female *slope* is significantly different than the male slope, and differences in adult-income means between the two groups will not matter.

5.2.1 Main results by race

Figure 6 shows Black and white mobility for the earlier, less-mobile 1910s–1920s cohorts compared to the later, more-mobile 1940s–1950s cohorts (IGE in panel (a) and rank-rank correlation in (b)). Perhaps the most striking aspect of the graph is how little overlap there is in the support of the Black versus white mobility graphs: Black fathers' income overlaps only modestly with white fathers' income. In the rank-rank figure, almost no white respondents grow up in the bottom ten percent of predicted family income and few Black respondents grow up with income above the 30$^{\text{th}}$ percentile, so the overlap of the two groups happens almost entirely between the tenth and thirtieth percentiles of parental income.[18]

[18] One feature of our "small data" is that the vast differences between how Black and white children grow up is readily apparent in the support of these figures: with full-population administrative data one can capture the tiny number of Black children who grew up in rich families and thus extend the regression lines over the entire 0–100 domain of parental income rank. But even today prime-age Black adults are

Another striking result is the significant progress Black respondents make relative to their white counterparts over this period. In the IGE graph, the entire Black regression line shifts upward by about twenty-five log points, whereas there is a much more modest upward shift for whites. The rank-rank graph shows a similar convergence, as we would expect. A Black child in the earlier cohorts growing up at the 25$^{\text{th}}$ percentile would be predicted to have an adult family income at the 30.8$^{\text{th}}$ percentile, compared to the 44.6$^{\text{th}}$ percentile for a similarly situated white child from this era. But for mid-century cohorts, Black children born at the 25$^{\text{th}}$ percentile are predicted to appear at the 37.7$^{\text{th}}$ percentile as adults compared to 45.8$^{\text{th}}$ for whites (nearly halving the gap with their white counterparts from around 14 to 8 percentile ranks).

While we have so far focused on Black-white convergence, the regression lines explaining white-only mobility also change over this period. In both the IGE and the rank-rank estimates, the slopes flatten significantly. The rank-rank slope falls from 0.28 to 0.20, the latter being close to the levels calculated in modern-day Canada or Denmark. As the large majority group, the flattening of the mobility slope among white individuals will have an important effect on the overall full-cohort IGE and rank-rank estimates, as is clear from the decomposition above.

5.2.2 Main results by gender

A major motivation for our family-income-to-occupation-score mobility concept is that it enables us to perform intergenerational-mobility estimation including women. As noted, the decomposition in equation (5) implies that including women will only affect the population-wide IGE if women's mobility *slopes* differs from that of men. *A priori*, there is no reason to assume that the mobility slopes of men and women will coincide. For example, marriage patters could differ by parental income and they will tend to matter more for women's family income than for men's, especially in the historical period when many married women did not work.

Figure 7 shows IGE and rank-rank estimates separately by gender. For both measures and for all birth decades, persistence measures for women are greater than or equal to those for men. The male-female mobility gap appears to be relatively stable over time, with perhaps a larger gap in the 1940s cohort.[19]

vastly under-represented in the upper parts of the parental income distribution while growing up. The tiny share of Black children in the upper ranks of parental income distribution even in modern data can be seen in the appendix figures of Chetty *et al.* (2020).

[19] As noted in Section 2, some of our datasets include only women (e.g., the National Longitudinal Surveys of Mature or Younger Women) or only men (the Occupational Changes in a Generation datasets), so a possible concern is that the differences in mobility by sex are an artifact of using different datasets. In Appendix Figure A.11 we show robustness to restricting the baseline sample to datasets that include both

Examining Figure 7 though the lens of equation (5), with the groups $g \in G$ being men and women, highlights that including women will tend to increase the overall IGE (and similarly, the rank-rank correlation). As women and men come from the same distributions of parental income (growing up together in the same households) and have generally equal population shares, the overall IGE is the simple average of the male and female IGE slopes. But since the pattern of IGE estimates for women closely track that of men (i.e., the male-female mobility gap is roughly steady) and women's population share will never stray far from fifty percent, it would be difficult to explain the *decline* in full-population intergenerational persistence with gender-specific factors.

While gender-specific factors are unlikely to explain the overall decline in persistence, it is still important to understand *why* women's adult family income depends more on their parents' income than is the case for men. To answer this question, we turn again to differences by race.

5.3 Main results by race and gender

We now consider differences by race separately for men and for women. In particular, Figures 8 and 9 further break down the by-race results in Figure 6 by gender. Figure 8 shows that among men, Black Americans closed almost the entire mobility gap with whites by mid-century. Of course, as the graph also makes clear by comparing the support of the predicted childhood income variable across the two distributions, Black men still grow up in far poorer households, so their adult income is still much lower than that of white individuals. But by mid-century, there is considerable overlap in adult outcomes between Black and white men born to similarly advantaged parents. Rank-rank relationships also become much more similar over time. In the more mobile mid-century cohorts, Black men born at the 25^{th} percentile are predicted to appear at the 41.3^{th} percentile as adults, just slightly below their white counterparts at the 45.8^{th} percentile. This 4.5 percentage point gap is 10.9 percentile points in the earlier cohorts, with Black men born at the 25^{th} percentile predicted to appear at the 34.3^{th} percentile as adults, compared to the 45.1^{th} percentile for their white counterparts.

A different story holds for Black women, as Figure 9 shows. First, comparing Figures 8 and 9, it is clear that Black adult women are simply poorer than their male counterparts. Their entire regression line is below that of men (whereas in both levels and slopes, white women have very similar mobility relationships to their male counterparts, again in both periods). In the early cohorts, a Black woman born at the 25^{th} percentile is predicted to barely climb upward at all (with an adult family-

men and women (roughly 47% of the baseline sample).

income percentile rank of 27.1). While Black women make progress over time, even at mid-century the corresponding prediction is only the 34.1^{th} percentile (the analogous statistic for white women is 45.8, the same as for white men). Thus, for mid-century cohorts, while the racial mobility gap at the 25^{th} percentile for men is down to 4.5 percentiles (from 10.9) it remains at 11.6 (down from 16.9) for women. Note that the lack of gender gaps by family income among white respondents and the large gaps (favoring men) among Black respondents is apparent in the basic summary statistics shown in Table 3, both in our surveys and in the Census.

As there appears to be almost no gender differences in adult outcomes among white respondents and because white men and women grow up in the same households, it follows that they likely have very similar mobility patterns. Indeed, comparing Figures 8 and 9 shows that the white-only mobility slopes are nearly identical for men and women. For the rank-rank correlation, the male and female slopes are both 0.28 in the early period and 0.20 and 0.21, respectively, in the later period. That white men and women's family incomes were equally tied to the status of their fathers in an era when most married white white women did not work suggests that they were marrying individuals very similar in earnings to their brothers.

These results highlight that the higher IGE and rank-rank persistence measures for women relative to men in Figure 7 are not driven by white individuals. Instead, the fact that Black women do poorly relative to Black men in adulthood pulls down the female mobility regression line for the lowest percentiles of parental income and results in a steeper slope for women relative to men throughout the first half of the twentieth century.

5.4 Comparing representative versus subgroup mobility estimates

In Figure 10, we show how the mobility estimates change as we sequentially add various subgroups. We begin with white men, the group most often studied in the existing mobility literature. The analysis above suggests that adding white women to this sample should not materially change the results, which is what we find in the second series of the figure. In some decades, adding women increases estimated persistence and in other decades it reduces it, but in all cases the deviations are modest.

We then add Black respondents, first men and then women. Both additions increase the estimated persistence measures in all decades, as we would expect from the evidence already presented. And, again as expected, the change tends to be larger once we add Black women. As they are born to families at the bottom of the distribution (like their male counterparts) and tend to remain poor as adults (more so than their male counterparts) adding this group can significantly change the slope of the full-population regression line, despite this group being just over five percent of the population. Indeed, their influence over the full-population regression line is readily apparent in the Figure.

In terms of the actual effects of using representative samples on various point estimates, take the 1920s cohort as an example. The white-male rank-rank slope is 0.27, which drops to 0.26 after adding white women. Adding Black men—just over five percent of the population—increases it an additional three percentage points to 0.29 and adding the similarly small group of Black women increases it to 0.32. Similarly, the IGE in for this cohort rises from 0.39 for white men to 0.47 for the representative population. Excluding Black men and especially Black women paints an overly *optimistic* picture about the *level* of intergenerational mobility in the first half of the twentieth century.

Considering a representative population instead of only white men also changes our view of the evolution of mobility over this period. For white men, the rank-rank correlation falls from 0.31 to 0.19 (a twelve-point decline) between the 1910s and 1940s cohorts. For the full-population, it falls from 0.39 to 0.24, a fifteen-point decline.[20] By missing out on Black-white convergence (which was especially large among men), examining only white men (or all whites) misses a substantial part of the decline in the slope and thus paints an overly *pessimistic* picture of the *rise* in intergenerational mobility over this same period.

5.5 Decomposing the decline in intergenerational persistence

As we already discussed, the full-population persistence slope is approximately equal to the (simple) mean of the male-only and female-only slopes. Because the gap between those two slopes is quite stable over time, a decomposition by sex is unlikely to help us explain the *decline* in full-population persistence over time. So we consider the decomposition by race instead. Returning to Figure 6 with the decomposition in mind allows us to assess the effects of the various movements in the by-race IGE and rank-rank mappings. Figure 6 depicts a number of different changes over time, some of which will increase mobility (the level increase for Black respondents, the slope decrease for white and Black individuals) and some of which will reduce mobility (the level increase for whites). The decomposition can quantify the various contributions.

We begin by considering the changes in levels—that is, a positive shift in real income for both Black and white cohorts over the first half of the twentieth century, though faster income growth for Black individuals, resulting in significant catch-up. Figure 11

[20] The analogous comparison with the IGE shows a nineteen-point decline when only looking at white men, and a 21-point decline when looking at the full population. Appendix Table A.2 reports estimates quantifying the differences between the 1910s and 1940s cohorts as well as from modeling the decline linearly for respondents born in the 1910s–1940s birth cohorts.

shows that if Black individuals had instead experienced the same real income growth as white individuals (without changing the slopes for either group), then 56 percent of the IGE and 45 percent of the rank-rank decline would not have been realized. Thus, Black respondents' catch-up to whites in income levels over this period explains a large share of the total decline in persistence, despite Black Americans only being a small share of the population.

We can also ask what share of the total decline in persistence is explained by the flattening of the white-only mobility slope. Given that white individuals are the majority group in the population, equation (5) suggests changes in their slope will matter for full-population mobility. Figure 11 shows that the change in the white slope accounts for much of the remaining change in mobility between the 1910s–1920s and 1940s–1950s cohorts, with very similar effects across the two measures (62% in the case of the IGE, 60% in the case of the rank-rank coefficient).[21]

6 Comparison to the modern period

Our focus has been on the first half of the twentieth century, given that our methodology does not allow us to reach back before the cohorts born in the 1910s and that for more modern cohorts other data are available, such as IRS administrative data. In this section, instead of focusing on the changes we document in our sample period, we consider how our results compare to the modern period.

While the comparison is not perfect given our different ways of calculating childhood household income, the results in Chetty *et al.* (2020) suggest a significant retrenchment of the progress Black Americans made during our sample period in closing the mobility gap with whites. For the 1978 to 1983 cohorts, they calculate a 12.5 percentile racial gap for children born at the 25th percentile of parental income. This result is similar to the 14-point gap we document for the 1910s–1920s cohorts and substantially larger than the eight-point gap we find for the 1940s–1950s cohorts.

The retrenchment of racial mobility gaps is sobering. While we found substantial convergence for the generation entering the labor market in the years of and immediately following the Civil Rights Era, much if not all of that convergence appears undone in the years since. The pattern of relative racial mobility improving over the 20th century but then reversing toward the end is perhaps unsurprising given similar patterns in the Black-white median income gaps (Bayer and Charles, 2018), and

[21] These two forces—the convergence of black-white mean income and the flattening of the white-only slope—in fact slightly over-explain the decline in persistence. Other factors—for example, the decline in the variance of parental income, which we document in Appendix Figure A.1—offset these two forces.

suggest a racial-inequality variant of the Great Gatsby curve: compression of racial differences in median income correlates with convergence in mobility differences between racial groups. If the reversion to pre-World-War-II levels of racial differences in mobility is a cause for pessimism, the mutability of mobility differences over the 20th century is a cause for optimism, suggesting that political and policy changes can indeed alter mobility patterns.

Examining changes in mobility by race *and* sex further support the idea that mobility patterns are sensitive to policy and institutional change. Figure 12 considers individuals growing up at the 25$^{\text{th}}$ percentile of the income distribution, and shows that it is the income growth of Black men that drives the Black-white convergence: those born at mid-century have nearly closed the mobility gap with white men, while Black women from these cohorts still have adult family income far below that of white women of similar childhood incomes.[22] While the modern data suggest that racial mobility gaps have regained their 1910–1920 levels, Chetty *et al.* (2020) show that the patterns today by sex are *reversed*.[23] Black men today significantly lag Black women not only in family income, but just about every observable measure of economic well-being (e.g., education, employment, individual earnings). Similarly, while we found essentially no gender gaps for whites over our sample period, white women today have a small but significant advantage over white men. That a gender gap that held for at least half a century can be reversed suggests to us that mobility patterns are not set in stone.

Looking forward, any attempt to raise intergenerational mobility for future generations must contend with the powerful countervailing force of between-group differences. Taking the summary statistics from the IRS data and our rank-rank decomposition in equation (4), we can see that today, roughly one-third of the full-population rank-rank coefficient of 0.36 that Chetty *et al.* (2020) estimate arises from between-racial-group differences in average income.[24] A striking result is that, for the U.S. full-population rank-rank correlation to fall to Danish or Canadian levels without between-group racial

[22] Recall our discussion of Appendix Figure A.2, where we show that beginning with the 1960s cohorts, a significant share of Black men ages 30–50 in Census data are institutionalized and thus unlikely to be responding to the types of surveys we use in our analysis. For this reason, our estimates of the average income at the 25$^{\text{th}}$ percentile for Black men are likely too high for the 1960s–1970s cohorts.

[23] Many other papers have shown that along many key dimensions, Black women outperform men, either in absolute terms or relative to the white gender gap. See, e.g., Autor *et al.* (2019) and papers cited therein.

[24] In this exercise, we use aggregate administrative data made public by the Opportunity Insights project. We use the average household income rank of adult children as well as the population distributions by parental income rank for six subgroups (individuals who are white, Black, Hispanic, Asian, and American Indian & Alaska Native as well as those persons for whom race is missing). We decompose the rank-rank correlation into the within-group and between-group components, and find that the latter component accounts for 28% of the overall measure.

means compressing, the within-group rank-rank would have to be less than 0.1, instead of over 0.25, as they are today.[25] Put differently, U.S. intergenerational persistence faces a high lower bound unless major convergence across racial groups occurs.

The comparison to modern data also suggest at least two areas for further work. First, marriage patterns will have a first-order effect on mobility when we consider representative cohorts and thus include women. The lower marriage rates of Black Americans relative to whites throughout our sample period (see Table 3) and continuing today permit large mobility gaps between men and women (as they are not married to each other and thus do not mechanically share a family income).[26] Interestingly, white marriage rates have also declined significantly in recent years, and a white gender mobility gap (in the same direction as for Black Americans, favoring women) has emerged in the modern data as well.[27] In addition to studying the implications of declining marriage rates for intergenerational mobility, future work might also examine the rise of interracial marriage—while the rates were negligible during our sample period, today 18 percent of recently married Black Americans have a spouse of a different race.[28]

Second, any candidate explanation for the reversal of Black progress in closing the mobility gap with whites would need to have a large gender-specific component, given the progress Black women have made. A natural candidate would be mass incarceration, a phenomenon that largely post-dates our historical cohorts but has important implications for modern cohorts of Black men. Chetty *et al.* (2020) document that approximately twelve percent of Black men who grew up at the 25$^{\text{th}}$ percentile of parental income were incarcerated on April 1st, 2010 (compared to less than four percent of white men). By contrast, for the 1940s and 1950s cohorts, important institutional changes *benefited* Black men economically (not just the employment protections from

[25] To simplify this calculation, we assume that all racial groups would have the same within-group rank-rank correlation. Chetty *et al.* (2020) show that within-group slopes are quite similar across groups (ranging from the mid-twenties to the low thirties), with the higher within-group mobility of Asian individuals being the exception.

[26] On face it may seem odd that in both our data and in the Census more Black men are married than Black women, especially given that there is essentially no interracial marriage in our historical period. However, past work suggests that there is significant misreporting of marital status, at least in the Census. While they look at a slightly earlier period than we do, Preston *et al.* (1992) argue that Black women often reported widowhood when in fact the husband had merely left (whereas the man might have reported still being married, just living separately), and that the Census generally under-reports marital instability for Black couples.

[27] The never-married share of all white adults ages 25 and over has doubled, from eight percent in 1960 to sixteen percent today. See `https://www.pewresearch.org/social-trends/2014/09/24/record-share-of-americans-have-never-married/`.

[28] See `https://www.pewresearch.org/social-trends/2017/05/18/1-trends-and-patterns-in-intermarriage`.

the Civil Rights movement, but outside the South a labor movement in manufacturing that by the 1960s was over-representing Black workers; see Farber *et al.*, 2021).

7 Conclusion

We provide, to the best of our knowledge, the first evidence on long-run intergenerational relative mobility trends for representative samples of the U.S.-born population. We find a robust decline in IGE and rank-rank persistence measures from the 1910s to the 1940s birth cohorts. For cohorts born after the 1940s, we find that the IGE—and to a lesser extent, the rank-rank—measure drifts back upward, thereby tracking the u-shape in the inequality measures of family income of the adult children in our sample. We thus provide some of the first evidence that the "Great Gatsby Curve" holds using within-country variation across cohorts, adding to existing work that finds this relationship using cross-sectional variation across countries in modern data.

Considering representative populations instead of just white men changes the overall view of mobility for cohorts born in the first half of the twentieth century. In any given year, mobility is lower in the full population than for white men alone. Including Black Americans (a group that is on average born to poor parents and who remain poor as adults) increases estimates of intergenerational persistence significantly, despite being only 10–12 percent of the population. In particular, Black women play an outsize role in increasing historical persistence measures because while they are born to equally poor parents as their male counterparts, they were, until recently, poorer as adults.

While looking only at white men in any given cohort biases mobility estimates downward relative to the full population, it also understates the rise in mobility from the 1910s to the 1940s (significantly so for the rank-rank coefficient). Including only white men misses out on the important progress Black Americans make relative to whites, which has large implications for full-population mobility given the extreme disadvantage of Black children over our sample period. In short, the United States starts the twentieth century much further from the "American Dream" ideal of a mobile society but also improves more significantly when the full population is considered rather than only white men.

The decline in intergenerational persistence over the 20th century we document challenges some of the recent and not-so-recent scholarship that has concluded that intergenerational mobility remains stable even in the face of large political and structural changes (see, e.g., Erikson and Goldthorpe, 2002, Clark, 2015, Ager *et al.*, 2019, Alesina *et al.*, 2020). Our cohorts span the mechanization and declining importance of American agriculture, the high school movement, two World Wars, the Great Depression, the New Deal, an unprecedented period of mass prosperity, and the Civil

Rights movement. Quantitatively unpacking the relative contributions of these forces in explaining the increase in mobility over the 20th century remains a fruitful area for future research.

References

Abramitzky, R., Boustan, L. P. and Eriksson, K. (2012). Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration. *American Economic Review*, **102** (5), 1832–56.

—, —, —, Feigenbaum, J. J. and Pérez, S. (2019). *Automated linking of historical data*. Tech. rep., National Bureau of Economic Research.

Ager, P., Boustan, L. P. and Eriksson, K. (2019). *The intergenerational effects of a large wealth shock: White southerners after the Civil War*. Tech. rep., National Bureau of Economic Research.

Alesina, A., Stantcheva, S. and Teso, E. (2018). Intergenerational mobility and preferences for redistribution. *American Economic Review*, **108** (2), 521–54.

Alesina, A. F., Seror, M., Yang, D. Y., You, Y. and Zeng, W. (2020). *Persistence Despite Revolutions*. Working Paper 27053, National Bureau of Economic Research.

Asher, S., Novosad, P. and Rafkin, C. (2018). Intergenerational mobility in india: Estimates from new methods and administrative data. *World Bank Working Paper*.

Autor, D., Figlio, D., Karbownik, K., Roth, J. and Wasserman, M. (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economic Journal: Applied Economics*, **11** (3), 338–81.

Bailey, M. J., Cole, C., Henderson, M. and Massey, C. (2020). How well do automated linking methods perform? lessons from u.s. historical data. *Journal of Economic Literature*, **58** (4), 997–1044.

Bayer, P. and Charles, K. K. (2018). Divergent paths: A new perspective on earnings differences between black and white men since 1940. *The Quarterly Journal of Economics*, **133** (3), 1459–1501.

Becker, G. and Tomes, N. (1979). An equilibrium theory of the distribution of income and integenerational mobility. *JPE*, **87** (6), 1153–1189.

Benabou, R. and Ok, E. A. (2001). Social mobility and the demand for redistribution: the poum hypothesis. *The Quarterly Journal of Economics*, **116** (2), 447–487.

Berman, Y. (2018). The long run evolution of absolute intergenerational mobility. *Available at SSRN 3269831*.

Bowles, S. (1972). Schooling and inequality from generation to generation. *Journal of Political Economy*, **80** (3, Part 2), S219–S251.

Card, D., Domnisoru, C. and Taylor, L. (2018). *The intergenerational transmission of human capital: Evidence from the golden age of upward mobility*. Tech. rep., National Bureau of Economic Research.

Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R. and Narang, J. (2017). The fading american dream: Trends in absolute income mobility since

1940. *Science*, **356** (6336), 398–406.

— and Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, **133** (3), 1107–1162.

— and — (2018b). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, **133** (3), 1163–1228.

—, —, Jones, M. R. and Porter, S. R. (2020). Race and economic opportunity in the united states: An intergenerational perspective. *The Quarterly Journal of Economics*, **135** (2), 711–783.

—, —, Kline, P. and Saez, E. (2014a). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The Quarterly Journal of Economics*, **129** (4), 1553–1623.

—, —, —, — and Turner, N. (2014b). Is the united states still a land of opportunity? recent trends in intergenerational mobility. *American Economic Review*, **104** (5), 141–47.

Clark, G. (2015). *The son also rises: Surnames and the history of social mobility*. Princeton University Press.

Collins, W. J. and Wanamaker, M. H. (2017). *African American Intergenerational Economic Mobility Since 1880*. Tech. rep., National Bureau of Economic Research.

Connolly, M., Corak, M. and Haeck, C. (2019). Intergenerational mobility between and within canada and the united states. *Journal of Labor Economics*, **37** (S2), S595–S641.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, **27** (3), 79–102.

Davis, J. and Mazumder, B. (2017). The decline in intergenerational mobility after 1980.

Erikson, R. and Goldthorpe, J. H. (2002). Intergenerational inequality: A sociological perspective. *Journal of Economic Perspectives*, **16** (3), 31–44.

Farber, H. S., Herbst, D., Kuziemko, I. and Naidu, S. (2021). Unions and inequality over the twentieth century: New evidence from survey data. *The Quarterly Journal of Economics*, **136** (3), 1325–1385.

Feigenbaum, J. J. (2015). Intergenerational mobility during the great depression.

— (2018). Multiple measures of historical intergenerational mobility: Iowa 1915 to 1940. *The Economic Journal*, **128** (612), F446–F481.

Ferrie, J. P. (1996). A new sample of males linked from the public use microdata sample of the 1850 us federal census of population to the 1860 us federal census manuscript schedules. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, **29** (4), 141–156.

Goldenweiser, E. A. (1916). The farmer's income. *The American Economic Review*, **6** (1), 42–48.

Haider, S. and Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, **96** (4), 1308–1320.

Heckman, J., Pinto, R. and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, **103** (6), 2052–86.

Helsø, A.-L. (2021). Intergenerational income mobility in denmark and the united states*. *The Scandinavian Journal of Economics*, **123** (2), 508–531.

Hertz, T. (2008). A group-specific measure of intergenerational persistence. *Economics Letters*, **100** (3), 415–417.

Hilger, N. G. (2015). *The great escape: Intergenerational mobility in the united states since 1940*. Tech. rep., National Bureau of Economic Research.

Hout, M. (1988). More universalism, less structural mobility: The american occupational structure in the 1980s. *American Journal of sociology*, **93** (6), 1358–1400.

Landersø, R. and Heckman, J. J. (2017). The scandinavian fantasy: The sources of intergenerational mobility in denmark and the us. *The Scandinavian journal of economics*, **119** (1), 178–230.

Long, J. and Ferrie, J. (2013). Intergenerational occupational mobility in great britain and the united states since 1850. *American Economic Review*, **103** (4), 1109–37.

Mazumder, B. (2015). Estimating the intergenerational elasticity and rank association in the us: Overcoming the current limitations of tax data.

— (2018). Intergenerational mobility in the united states: What we have learned from the psid. *The Annals of the American Academy of Political and Social Science*, **680** (1), 213–234.

Merriam, W. R. (1902). *Census Reports: Twelfth Census of the United States, Taken in the Year 1900*, vol. 2. Washington, United States Census Office.

Mogstad, M. and Torsvik, G. (2021). Family background, neighborhoods and intergenerational mobility. Working Paper.

O'Hare, W. P. (2019). Census coverage of the black population. In *Differential Undercounts in the US Census*, Springer, pp. 83–91.

Olivetti, C. and Paserman, M. D. (2015). In the name of the son (and the daughter): Intergenerational mobility in the united states, 1850-1940. *American Economic Review*, **105** (8), 2695–2724.

Preston, S. H., Elo, I. T., Hill, M. E. and Rosenwaike, I. (2003). *The demography of African Americans 1930–1990*. Springer Science & Business Media.

— and Haines, M. R. (1991). Appendix a, assigning income and unemployment estimates to individuals in the national sample of the 1900 united states census. In *Fatal Years: Child Mortality in Late Nineteenth-Century America*, Princeton University Press, pp. 211–220.

—, Lim, S. and Morgan, S. P. (1992). African-american marriage in 1910: Beneath the surface of census data. *Demography*, **29** (1), 1–15.

Roemer, J. E. and Trannoy, A. (2015). Equality of opportunity. In *Handbook of income distribution*, vol. 2, Elsevier, pp. 217–300.

Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M. and Sobek, M. (2021). IPUMS USA: Version 11.0 [dataset].

Schoeni, R. F. and Wiemers, E. E. (2015). The implications of selective attrition for estimates of intergenerational elasticity of family income. *The Journal of Economic Inequality*, **13** (3), 351–372.

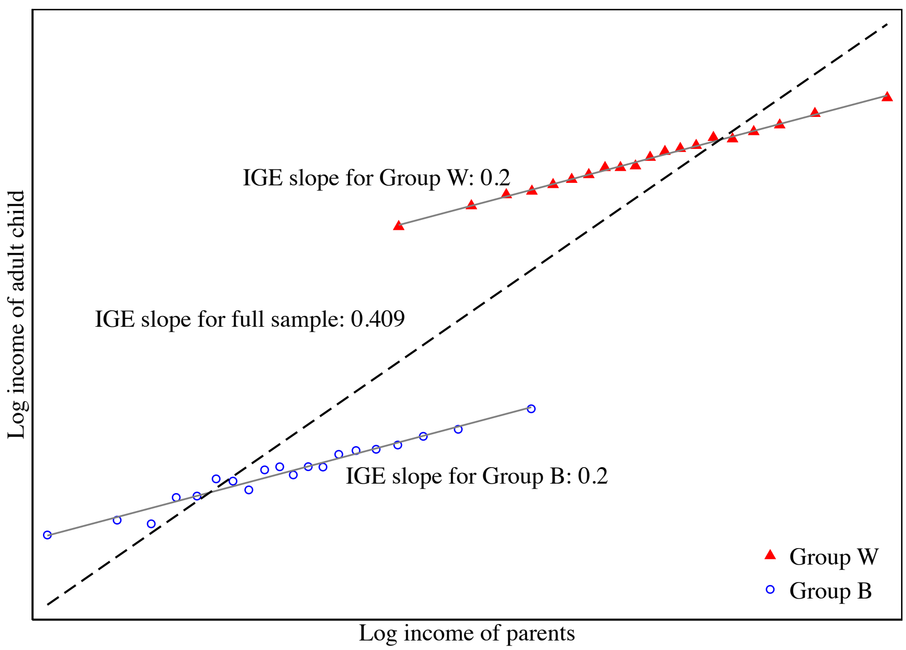
Solon, G. (1992). Intergenerational income mobility in the united states. *American Economic Review*, **82** (3), 393–408.

Song, X., Massey, C. G., Rolf, K. A., Ferrie, J. P., Rothbaum, J. L. and Xie, Y. (2020). Long-term decline in intergenerational mobility in the united states since the 1850s. *Proceedings of the National Academy of Sciences*, **117** (1), 251–258.

Uguccioni, J. (2021). Is timing everything? parental unemployment and their children's future earnings, mimeo, University of Toronto.

Ward, Z. (2020). Intergenerational mobility in american history: Accounting for race and measurement error.

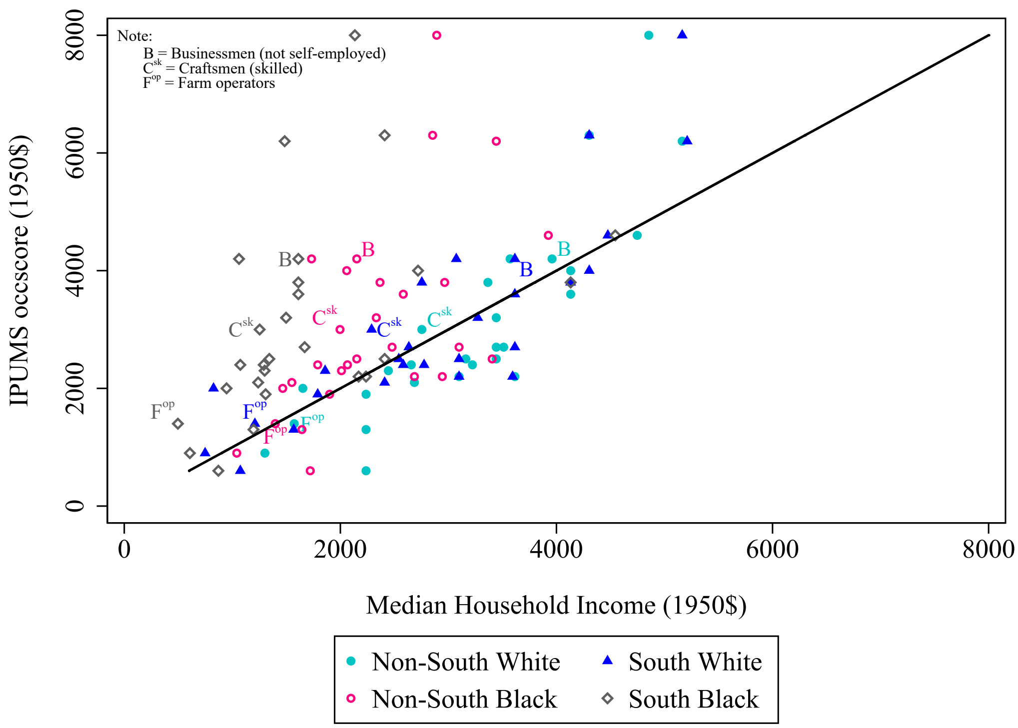
Figure 1: Illustration that IGE slope is not a weighted average of sub-sample IGE slopes



Sources: Data generated by the authors for the sake of illustration.

Notes: In this case, group W is the large majority (90% of the population) and B is the minority (10% of the population).

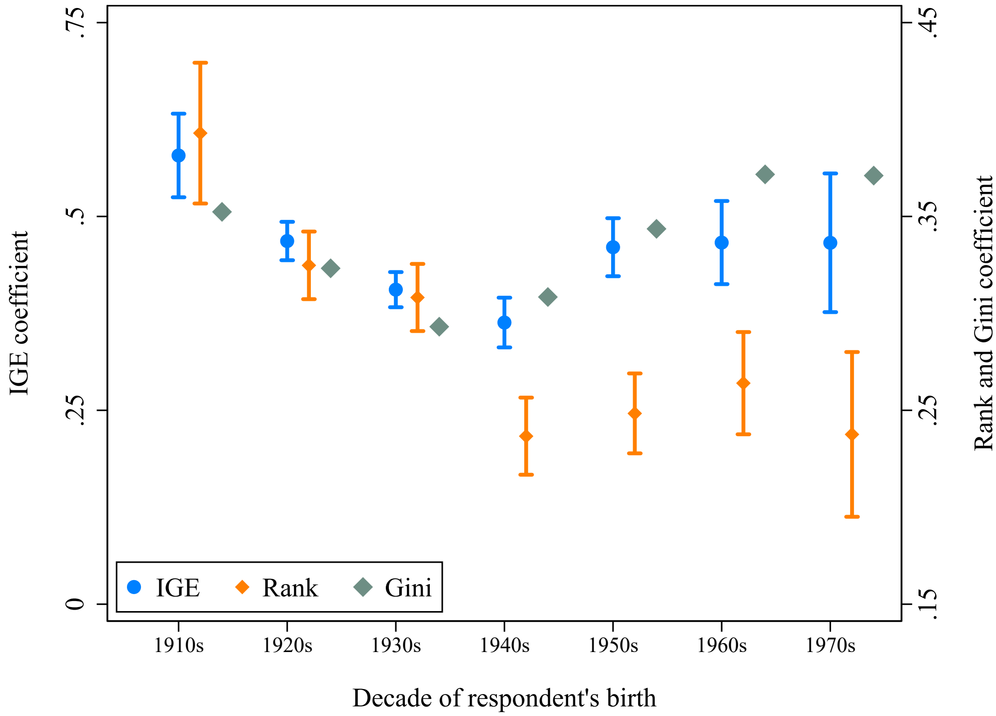
Figure 2: Comparing our income scores to the IPUMS *occscore* variable



Sources: 1940 and 1960 Censuses.

Notes: These income scores reflect the income of men between the ages of 30–50 living with at least one biological child under the age of 18. The y-axis corresponds to the income score using the *occscore* variable generated by IPUMS. The x-axis corresponds to our baseline income score (which is calculated specifically for this sub-sample of fathers and predicts income as a function of occupation, race, and Southern residence). We include a 45-degree line to aid comparisons. We highlight with labels a few salient occupations.

Figure 3: IGE and rank-rank measures by birth decade

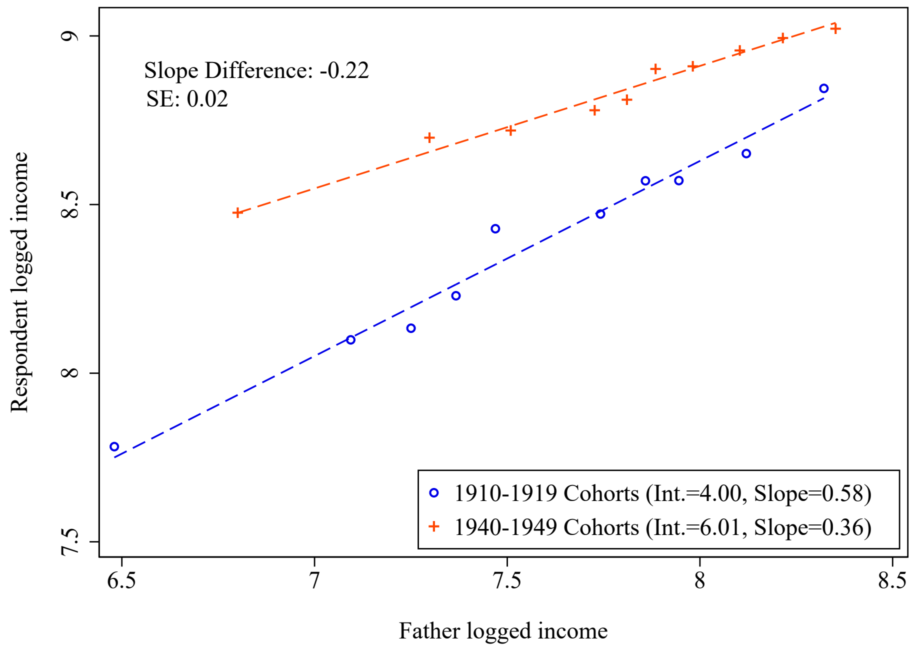


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

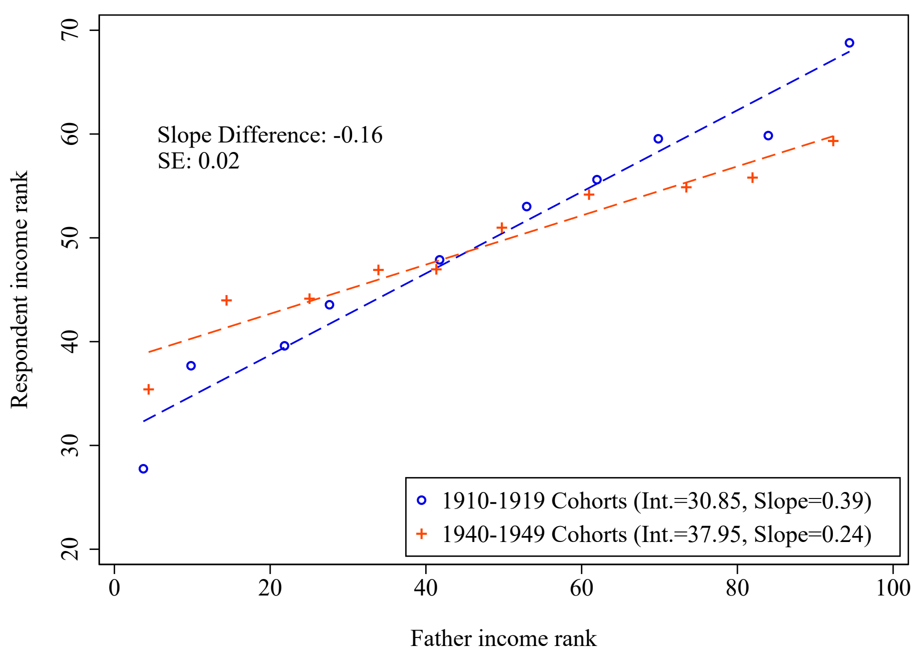
Notes: The IGE and rank-rank estimates are based on the baseline sample of respondents ages 30–50. The Gini coefficient uses the self-reported family income of the adult children (i.e., the respondents in our surveys). We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* $\times$ *sex* shares.

Figure 4: Bin-scatter depictions of the decline in intergenerational persistence

(a) Intergenerational elasticities



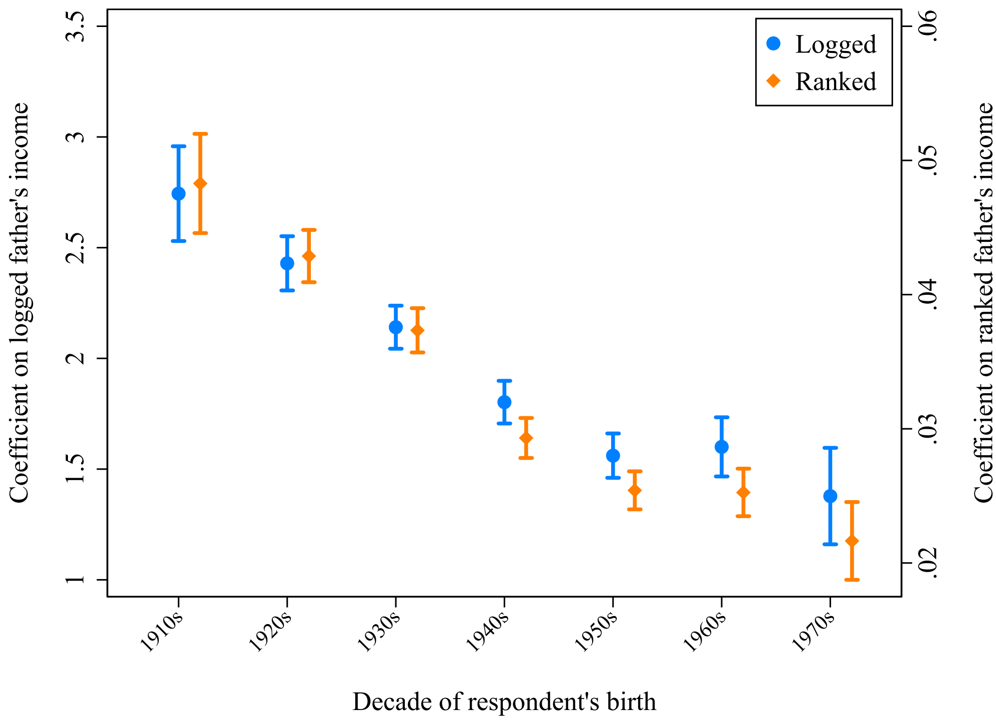
(b) Rank-rank relationships



Sources: Data come from 15 different surveys, described in Section 2 and in further detail in Appendix D.

Notes: The IGE and rank-rank are based on the baseline sample of respondents ages 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* $\times$ *sex* shares. The estimated slope difference and its standard error come from regressions similar to equations (1) and (2), but which allow the slope and intercept to differ by cohort.

Figure 5: Weakening relationship between respondent's educational attainment and father's income, by birth cohort

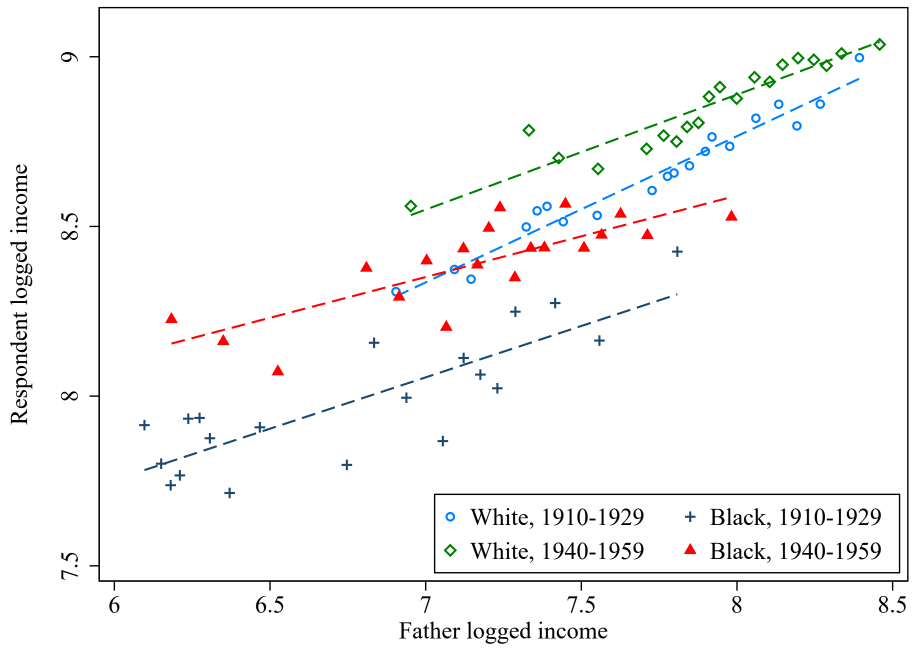


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

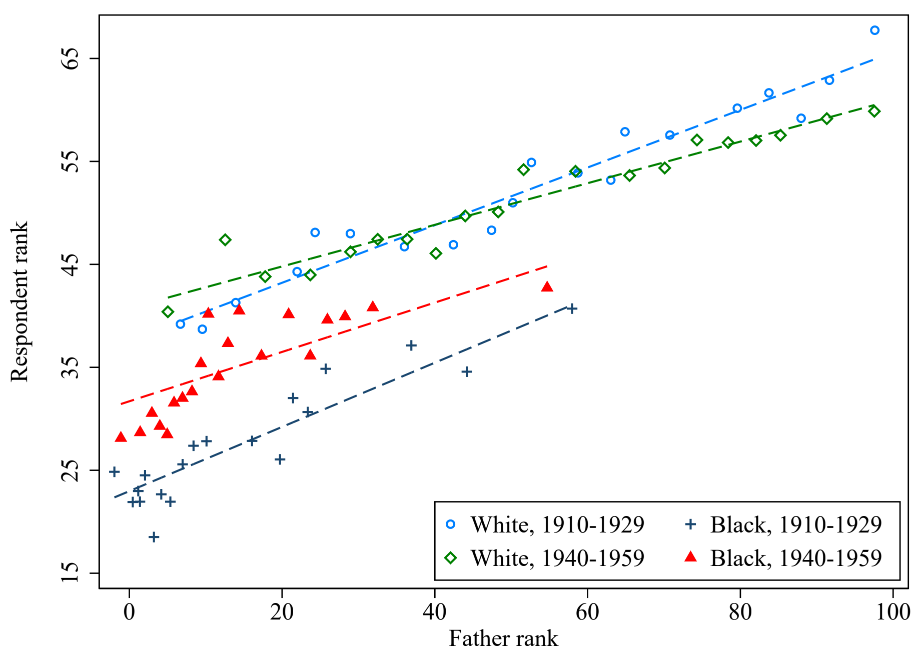
Notes: The IGE and rank-rank are based on the baseline sample of respondents ages 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* $\times$ *sex* shares. We use a respondent's years of schooling as the dependent variable and regress it on logged or ranked father's income score, similar to equations (1) and (2).

Figure 6: Mobility by race, 1910s–1920s versus 1940s–1950s

(a) Intergenerational elasticities



(b) Rank-rank relationships

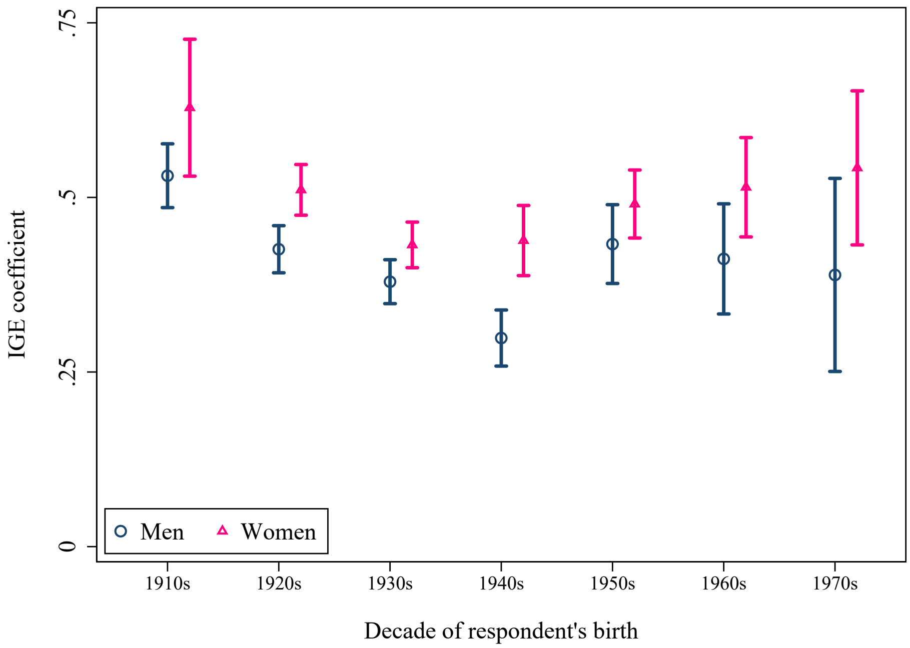


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

Notes: The IGE and rank-rank are based on the baseline sample of respondents ages 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

Figure 7: IGE and rank-rank measures by birth decade, by sex

(a) Intergenerational elasticity



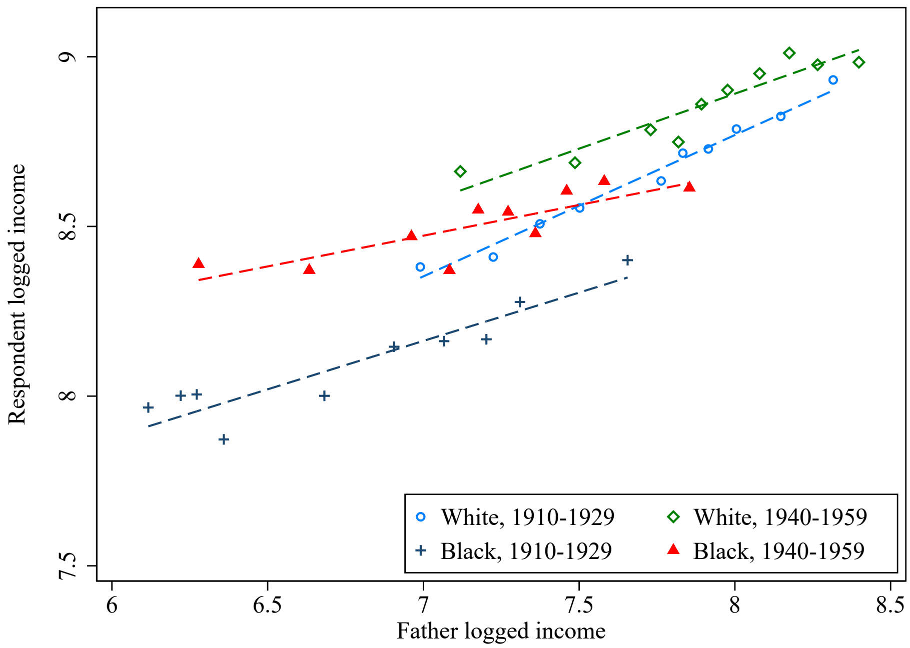
(b) Rank-rank coefficient

Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

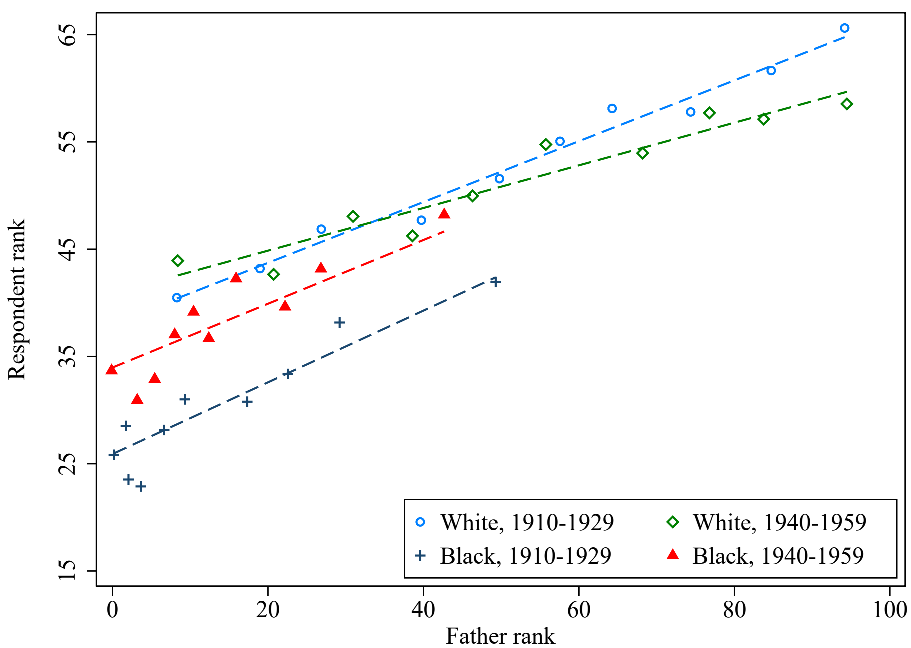
Notes: The IGE and rank-rank are based on the baseline sample of respondents ages 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

Figure 8: Mobility by race for men, 1910s–1920s versus 1940s-1950s

(a) Intergenerational elasticities



(b) Rank-rank relationships

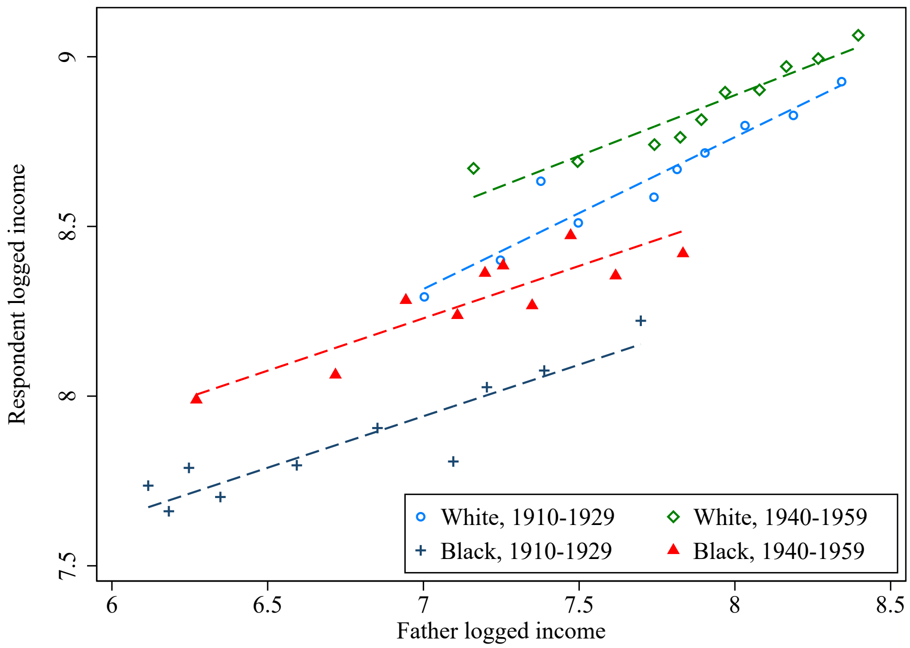


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

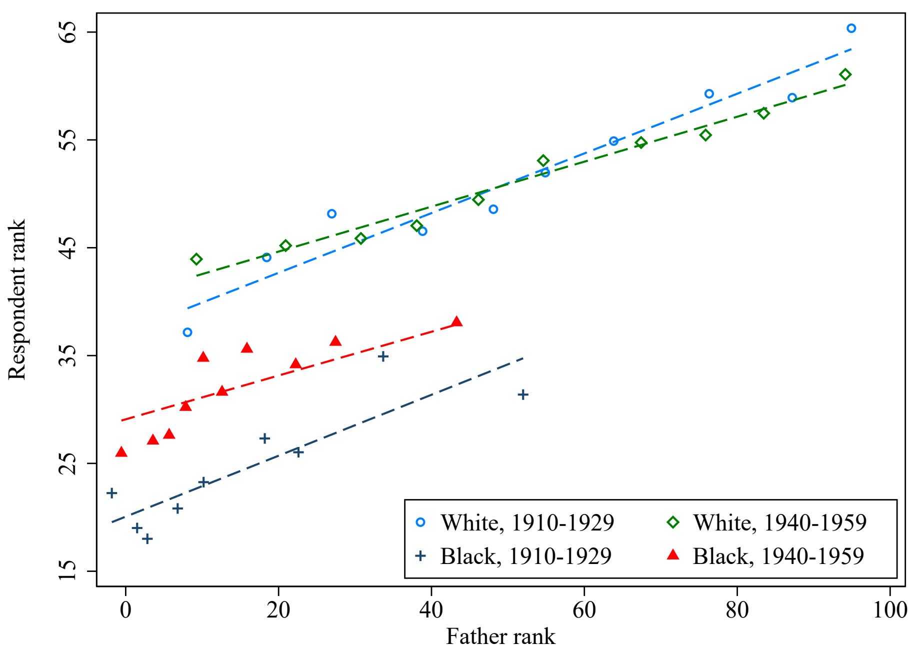
Notes: The IGE and rank-rank are based on the baseline sample of respondents ages 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

Figure 9: Mobility by race for women, 1910s–1920s versus 1940s-1950s

(a) Intergenerational elasticities



(b) Rank-rank relationships

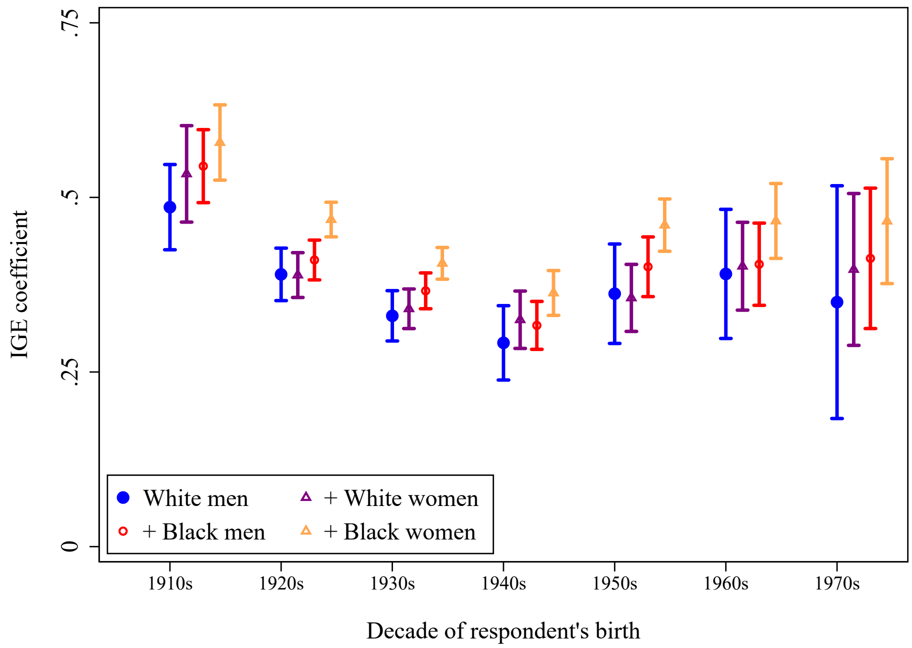


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

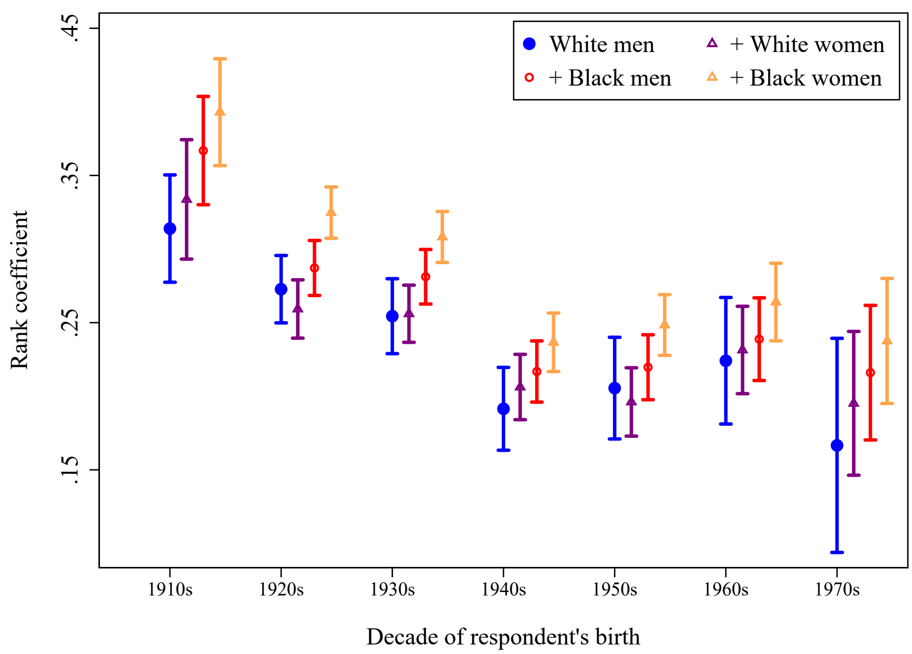
Notes: The IGE and rank-rank are based on the baseline sample of respondents ages 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

Figure 10: Mobility patterns over the 20th century including under-represented groups

(a) Intergenerational elasticities



(b) Rank-rank relationships

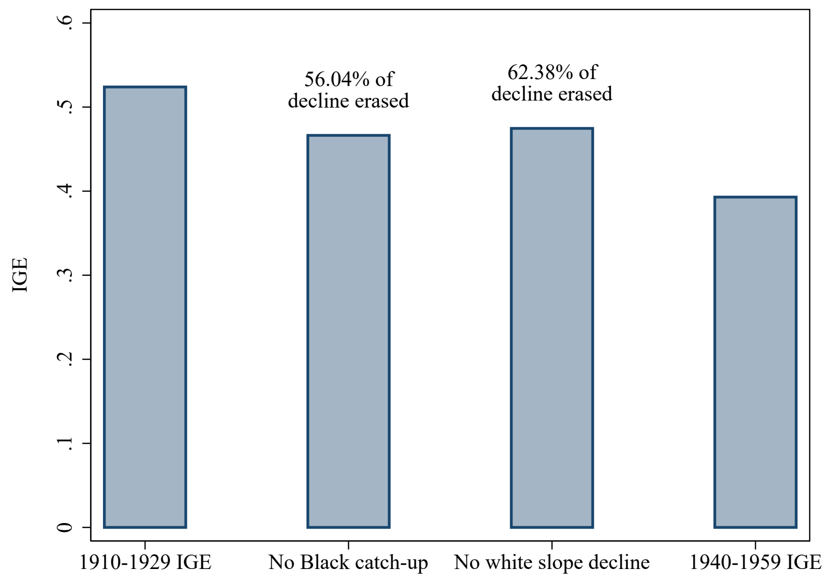


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

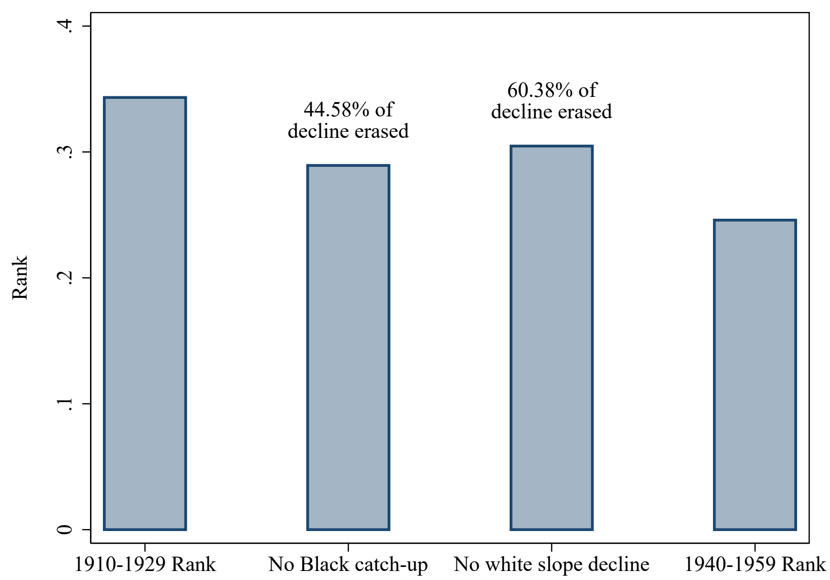
Notes: The IGE and rank-rank are based on the baseline sample of respondents ages 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

Figure 11: Decomposing the rise in mobility from the 1910s–1920s to 1940s–1950s

(a) Decomposition of IGE



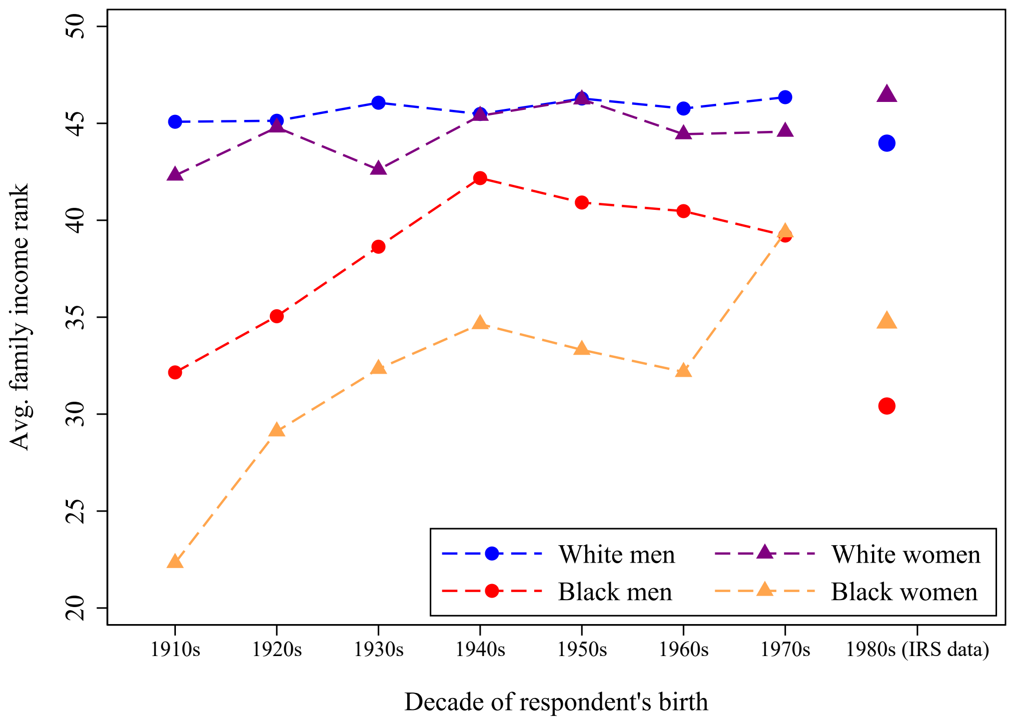
(b) Decomposition of Rank-Rank



Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

Notes: This figure shows the contribution of different components of the decomposition in Section 5 to the change in intergenerational mobility that occurred between the cohorts born in 1910–1929 and those born in 1940–1959. It shows the contribution of the reduction in the white-only IGE and the contribution of the between-group convergence in income levels. In the top panel, "No Black catch-up" refers to assuming that Black respondents had the same income growth as white respondents in log points. In the bottom panel, "No Black catch-up" refers to assuming that Black respondents in the later cohorts had the same average rank in adulthood as Black respondents in the earlier cohorts. In both panels, "No white slope decline" refers to white individuals in the 1940–1959 birth cohorts having the same slope as the 1910–1929 cohorts.

Figure 12: Average income rank of individuals born to the 25$^{\text{th}}$ percentile of the parental income distribution, by subgroup and birth cohort



Sources: Data for the 1910s–1970s birth cohorts combine 15 different surveys, which are described in Section 2 and in further detail in Appendix D. Data for the 1980 birth cohort come from Chetty *et al.* (2020), available at `https://opportunityinsights.org/data/`.

Notes: This figure plots plots the average adult income rank for individuals growing up at the 25$^{\text{th}}$ percentile of the parental income distribution, separately by race, sex, and birth cohort. For survey respondents, we use equation (2) to compute the expected income rank for individuals growing up at the 25$^{\text{th}}$ percentile of the parental income distribution. For the 1980 cohort, we use the average percentile rank in the national distribution of household income (measured in 2014–2015) for individuals growing up at the 25$^{\text{th}}$ percentile of the parent household income distribution (measured in 1994–2000).

Table 1: Select review of intergenerational mobility papers using U.S. data

Paper	Cohorts	Income/status proxy: Parent(s)	Income/status proxy: Child	Links	Sample
Ward (2020)	1850–1910	*Occ.*×*Race*×*Region*	Occ.	Match	All ♂
Collins and Wanamaker (2017)	1880–1970	*Occ.*×*Race*×*Region*	Occ.	Match & Retr.	All ♂
Song *et al.* (2020)	1830–1980	Occ.	Occ.	Match & Retr.	White ♂
Long and Ferrie (2013)	1840, 1930	Occ.	Occ.	Match & Retr.	White ♂
Olivetti and Paserman (2015)	1840–1910	Occ.	Occ.	Synthetic panel	White ♂ & married ♀
Feigenbaum (2018)	1900	Inc.	Inc.	Match	Iowa ♂
Feigenbaum (2015)	1900–1910	Inc.	Inc.	Match	Urban ♂
Card *et al.* (2018)	1920	Edu.	Edu.	Same household	Representative
Bowles (1972)	1930	Inc.	Inc.	Retrospective	CPS ♂
Mazumder (2015)	1950–1970	Inc.	Inc.	Panel data	Representative
Chetty *et al.* (2014a)	1980–1982	Inc.	Inc.	Claim dep.	Representative
Chetty *et al.* (2020)	1978–1983	Inc.	Inc.	Claim dep.	Representative
Our paper	1910–1970	*Occ.*×*Race*×*South*	Inc.	Retrospective	Representative

Notes: Since many papers do not explicitly consider birth cohorts, the "cohorts" column refers to the birth decade(s) that most of the sample comes from, given the age restrictions used in the paper. In the "Links" column, "Match" refers to matching across datasets (e.g., Census matching by name, age and state of birth); "Synthetic panel" refers to matching based on characteristics but not individual identity; "Claim dep." refers to matching by whether the parent ever claims the child as a dependent to the IRS; "Retrospective" (or "Retr.") refers to adult children being asked retrospectively about the characteristics of their parents (e.g., occupation and education).

Table 2: Summary statistics, by birth decade

	1910s	1920s	1930s	1940s	1950s	1960s	1970s
Father demographics:							
Foreign-born	0.22	0.17	0.11	0.05	0.03	0.03	0.05
High school educated	0.17	0.19	0.26	0.45	0.60	0.71	0.82
College educated	0.04	0.04	0.06	0.09	0.16	0.20	0.26
Farming occupation	0.37	0.29	0.24	0.15	0.09	0.05	0.03
Respondent demographics:							
Female	0.12	0.33	0.45	0.44	0.57	0.52	0.56
Age	45.89	41.52	36.95	38.49	38.05	38.47	38.92
Black	0.12	0.13	0.15	0.15	0.18	0.16	0.24
High school educated	0.50	0.61	0.71	0.85	0.90	0.91	0.91
College educated	0.10	0.14	0.16	0.28	0.28	0.30	0.39
Moved regions	0.21	0.21	0.22	0.24	0.22	0.21	0.22
Union member (men)	0.31	0.31	0.29	0.28	0.22	0.17	0.13
Veteran (men)	—	0.77	0.59	0.46	0.21	0.15	0.14
Father income:							
Income score (1950$)	2,171	2,240	2,309	2,533	2,693	2,785	2,699
Missing income	0.13	0.15	0.16	0.15	0.16	0.24	0.22
Father income rank	46.15	46.47	46.18	46.20	45.14	45.48	42.65
Respondent income:							
Family income (1950$)	5,506	6,803	7,292	7,896	7,620	7,891	8,483
Missing income	0.15	0.10	0.06	0.06	0.08	0.08	0.04
Bottom coded	0.08	0.05	0.03	0.02	0.04	0.05	0.06
Top coded	0.07	0.07	0.08	0.14	0.12	0.09	0.08
Family income rank	49.54	48.71	47.30	46.44	46.36	47.25	45.62
Observations	5,207	13,328	12,446	11,575	10,962	6,598	3,125

Notes: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D. All of the shares in this table are unweighted and are based on the baseline sample of respondents ages 30–50 (i.e., those with non-missing family income and childhood household income score). The two exceptions are the "Missing income" rows, which consider all U.S.-born respondents ages 30–50 in the 15 surveys. For characteristics that are unavailable in every survey (e.g., father's educational attainment), the average is computed using only the baseline-sample respondents in the surveys with the available information. When considering union membership and veteran status, we restrict the sample to male respondents. "Bottom coded" and "Top coded" refers to the share of individuals that had family income values in the bottom or top bins, respectively.

Table 3: Summary statistics, comparing survey respondents to Census respondents

	1910–1929		1930–1949		1950–1969	
	Census	Survey	Census	Survey	Census	Survey
Panel A: White Men						
Share of Men	0.90	0.91	0.90	0.90	0.87	0.86
Age	39.51	43.30	38.69	37.05	40.59	38.13
High school graduate	0.51	0.60	0.81	0.80	0.92	0.91
College graduate	0.12	0.15	0.27	0.27	0.30	0.33
Southern born/grew up	0.30	0.28	0.31	0.31	0.28	0.27
Resides in the South	0.28	0.27	0.33	0.31	0.34	0.31
Married	0.87	0.90	0.81	0.84	0.68	0.66
Widowed	0.01	0.01	0.00	0.01	0.00	0.01
Family income, 1950$	6,124	6,762	7,712	8,145	8,519	8,363
Respondent rank	52.57	51.64	53.20	52.60	52.20	53.49
Observations	195,091	12,281	214,612	11,939	297,783	6,742
Panel B: Black Men						
Share of Men	0.10	0.09	0.10	0.10	0.13	0.14
Age	39.41	44.57	38.54	37.66	40.13	38.02
High school graduate	0.21	0.28	0.62	0.60	0.85	0.82
College graduate	0.03	0.04	0.10	0.12	0.13	0.14
Southern born/grew up	0.86	0.84	0.77	0.73	0.60	0.62
Resides in the South	0.54	0.54	0.51	0.57	0.57	0.61
Married	0.75	0.82	0.63	0.69	0.50	0.53
Widowed	0.02	0.02	0.01	0.02	0.01	0.01
Family income, 1950$	3,817	4,257	5,738	6,109	6,318	6,134
Respondent rank	27.59	31.53	39.19	38.09	38.72	40.18
Observations	21,002	1,212	24,293	1,393	38,206	1,126
Panel C: White Women						
Share of Women	0.89	0.79	0.88	0.80	0.86	0.81
Age	39.50	40.97	38.74	38.56	40.64	38.35
High school graduate	0.55	0.66	0.81	0.82	0.94	0.93
College graduate	0.07	0.09	0.17	0.19	0.30	0.31
Southern born/grew up	0.30	0.31	0.31	0.31	0.28	0.27
Resides in the South	0.28	0.30	0.32	0.32	0.34	0.31
Married	0.86	0.86	0.79	0.77	0.70	0.65
Widowed	0.03	0.03	0.02	0.02	0.01	0.01
Family income, 1950$	6,033	6,865	7,527	7,737	8,469	8,061
Respondent rank	51.06	51.64	51.45	50.73	51.75	51.60
Observations	201,503	3,977	217,061	8,535	302,610	7,810
Panel D: Black Women						
Share of Women	0.11	0.21	0.12	0.20	0.14	0.19
Age	39.27	40.88	38.70	37.81	40.18	38.01
High school graduate	0.25	0.32	0.63	0.59	0.88	0.83
College graduate	0.04	0.05	0.09	0.11	0.17	0.16
Southern born/grew up	0.86	0.84	0.77	0.73	0.61	0.66
Resides in the South	0.55	0.60	0.51	0.57	0.58	0.64
Married	0.66	0.64	0.50	0.52	0.40	0.37
Widowed	0.08	0.09	0.06	0.06	0.03	0.03
Family income, 1950$	3,560	3,598	4,962	4,806	5,706	4,968
Respondent rank	23.72	23.81	32.87	29.15	34.65	32.52
Observations	24,081	1,065	29,808	2,154	45,166	1,882

Notes: Survey shares are based on the baseline sample of respondents ages 30–50 and are unweighted. We use the 1% samples of the 1960, 1980, and 2000 Censuses from Ruggles *et al.* (2021) and keep Census respondents born in the same years as survey respondents.

Table 4: IGE and rank coefficient, by birth cohort

(a) Intergenerational elasticity

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
IGE coefficient	0.579	0.468	0.406	0.363	0.460	0.466	0.466
	[0.028]	[0.013]	[0.012]	[0.016]	[0.019]	[0.027]	[0.046]
Observations	5,207	13,328	12,446	11,575	10,962	6,598	3,125

(b) Rank-rank coefficient

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
Rank coefficient	0.393	0.325	0.308	0.237	0.248	0.264	0.238
	[0.019]	[0.009]	[0.009]	[0.010]	[0.011]	[0.013]	[0.022]
Observations	5,207	13,328	12,446	11,575	10,962	6,598	3,125

Notes: The IGE and rank-rank estimates—calculated using equations (1) and (2), respectively—are based on the baseline sample of respondents ages 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

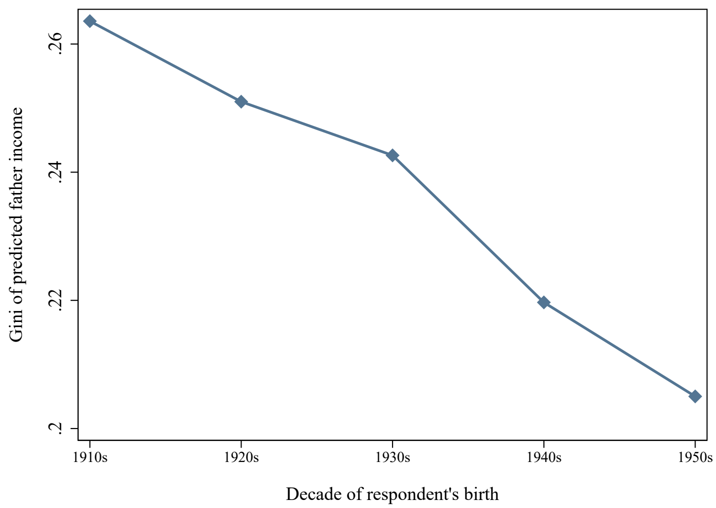
Mobility for All: Appendix Materials

List of Appendix Figures

List of Tables

A Additional figures and tables referenced in the text

Figure A.1: Declining Gini coefficient of predicted father's income, by birth cohort

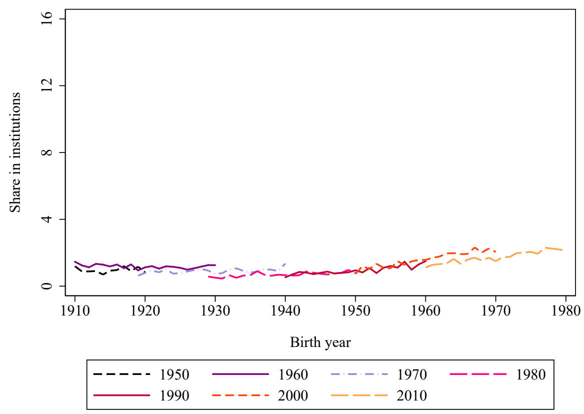


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

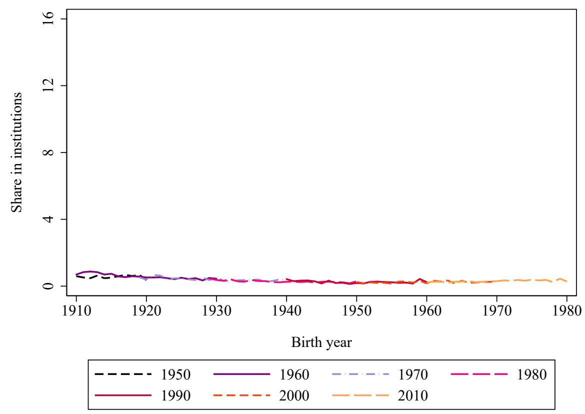
Notes: This figure plots the Gini coefficient of fathers' income scores, separately for each birth cohort in the baseline sample.

Figure A.2: Share of Census respondents in institutions, by birth year and Census year

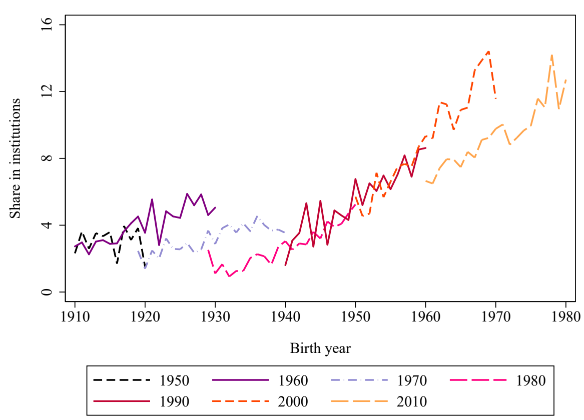
(a) White men



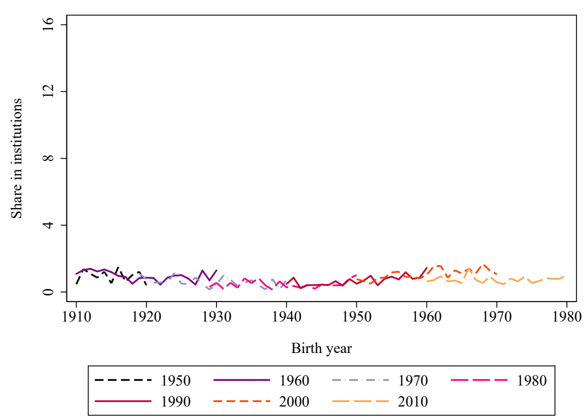
(b) White women



(c) Black men



(d) Black women

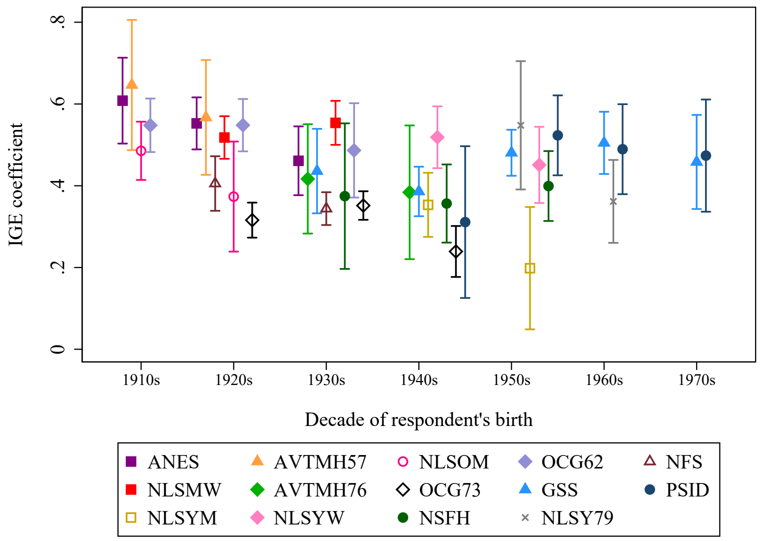


Sources: 1950–2000 1% Census samples and 2010 American Community Survey (Ruggles *et al.*, 2021).

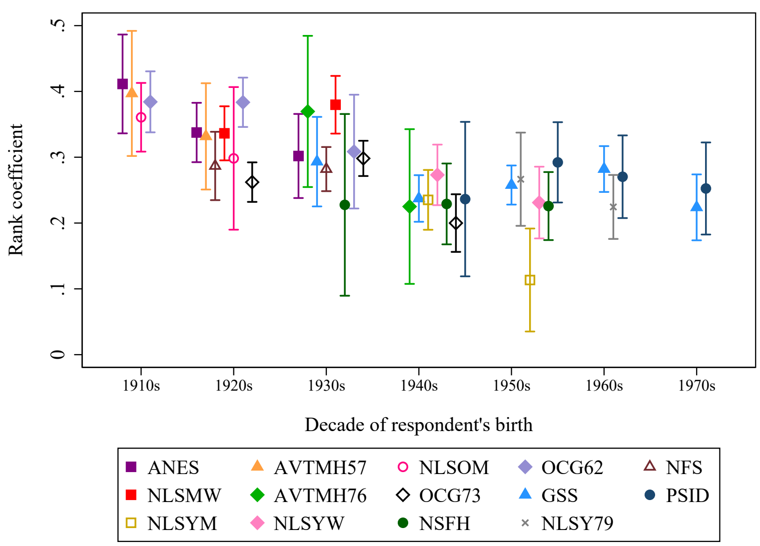
Notes: This figure plots the share of individuals born in a specific year that are living in institutions (measured using group quarter status), separately by Census year. The sample is restricted to white and Black U.S.-born Census respondents.

Figure A.3: Mobility measures by birth decade and by survey

(a) Intergenerational elasticity



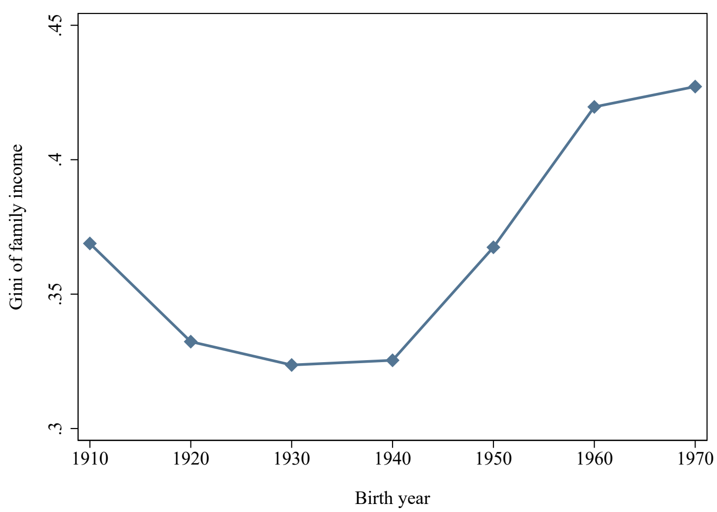
(b) Rank-rank coefficient



Sources: This figure uses 14 of the 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

Notes: Using the baseline sample and income scores, this figure plots the IGE and rank-rank coefficients estimated on each survey separately. We maintain the same ranks and weights as in our baseline specification. We exclude surveys whose respondents are all Black Americans as well as cohorts within a survey if there were fewer than 200 respondents born in that decade.

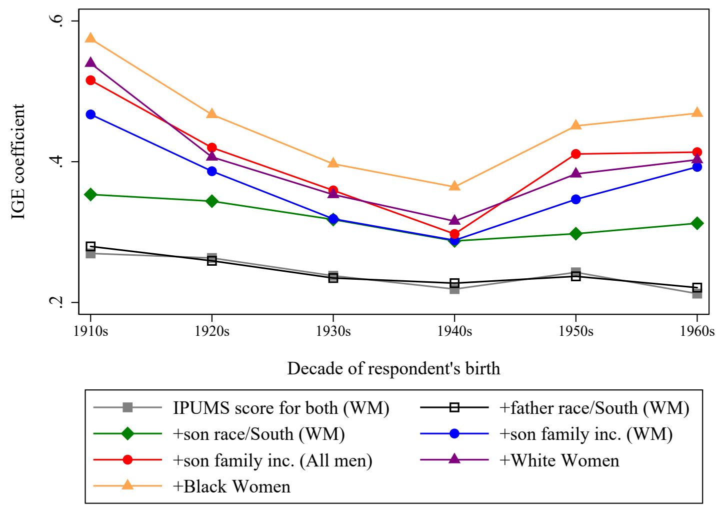
Figure A.4: Gini coefficient of adult family income in the Census, by birth cohort



Sources: 1950–2000 1% Census samples and 2010 American Community Survey (Ruggles *et al.*, 2021).

Notes: This figure plots the Gini coefficient of family income for individuals who are age 40 in each Census. The sample is restricted to white and Black U.S.-born Census respondents. Only one individual is counted per family, and we use total family income to calculate the Gini coefficient. The small share of families with negative or zero income is assigned an income of one so that they are included in calculations.

Figure A.5: IGE measure by birth cohort, using various samples and ways of measuring fathers' and children's incomes

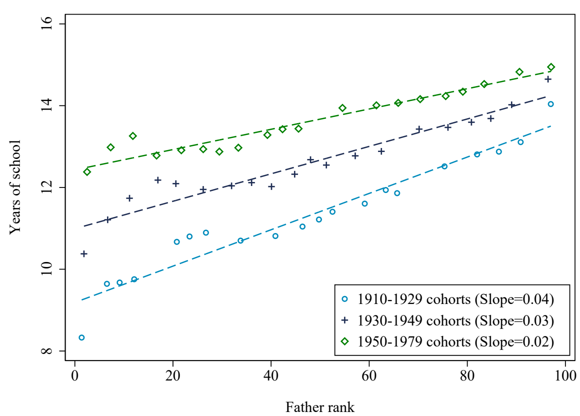


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

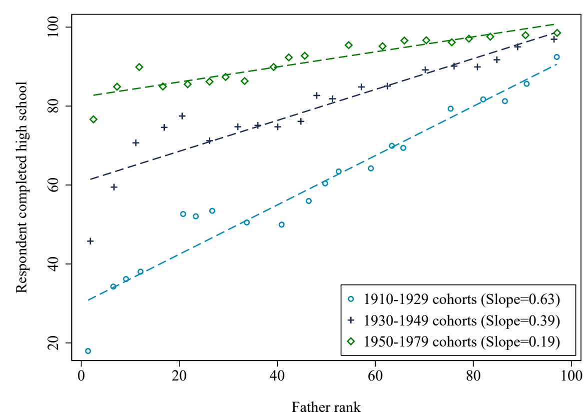
Notes: This figure plots the IGE coefficient as we alter how we measure respondents' and their fathers' incomes as well as vary the sample. The first series uses the IPUMS *occscore* variable for white male respondents who reported their occupation as well as for these respondents' fathers. The second series uses the same sample as the first series, but replaces the income scores of fathers with the baseline income scores, which vary at the *occupation* × *race* × *South* level. Maintaining the same sample, the third series uses the baseline income scores for both fathers and sons. The fourth series replaces the son's income score with the son's reported family income. The fifth, sixth, and seventh series all use the baseline measures of income for respondents and fathers (i.e., similar to the fourth series), but successively add Black men, white women, and Black women.

Figure A.6: Bin-scatter depictions of the weakening relationship between respondent education and father's income

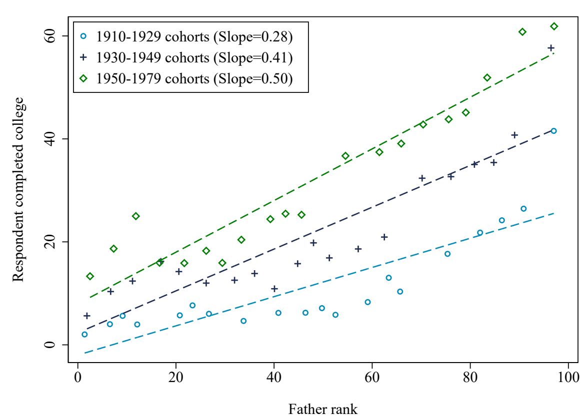
(a) Years of schooling



(b) High school completion



(c) College completion

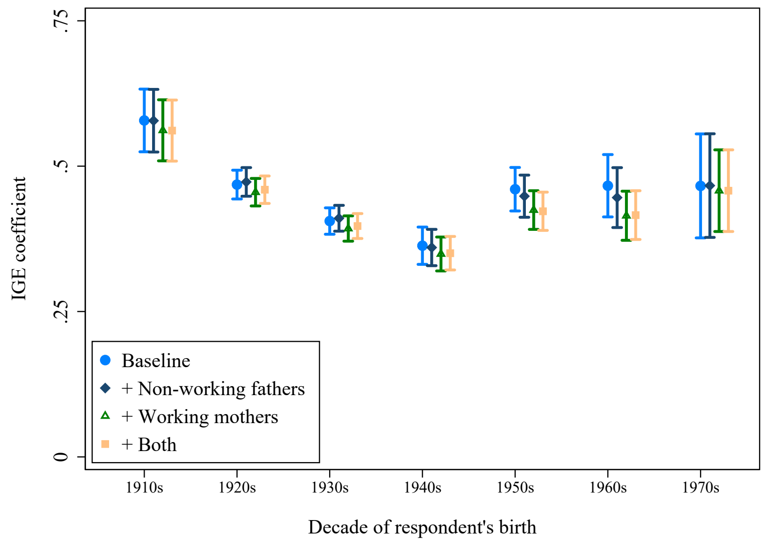


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

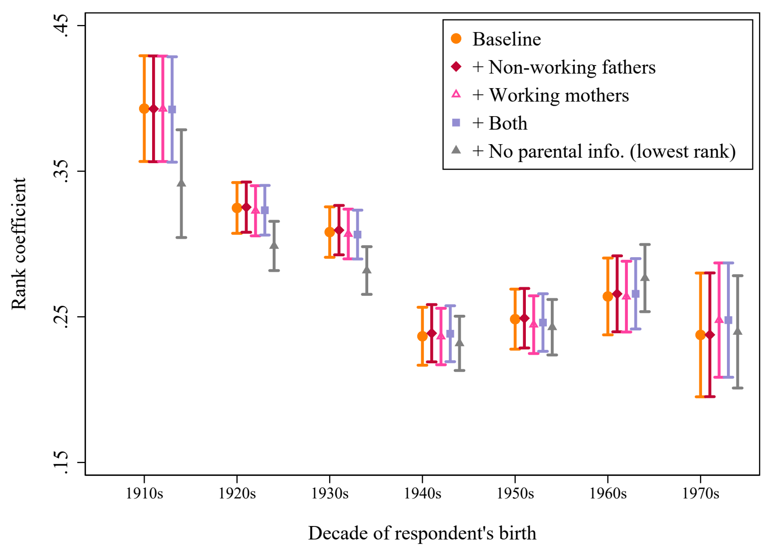
Notes: The estimates are based on the baseline sample of respondents ages 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* $\times$ *sex* shares. Further details on the construction of education variables are available in Appendix D.

Figure A.7: Mobility measures by birth decade, incorporating respondents with missing father's income

(a) Intergenerational elasticity



(b) Rank-rank coefficient

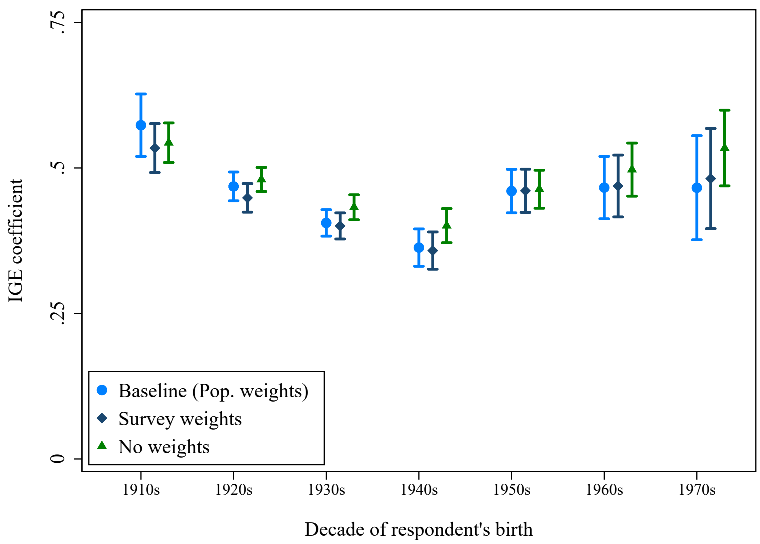


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

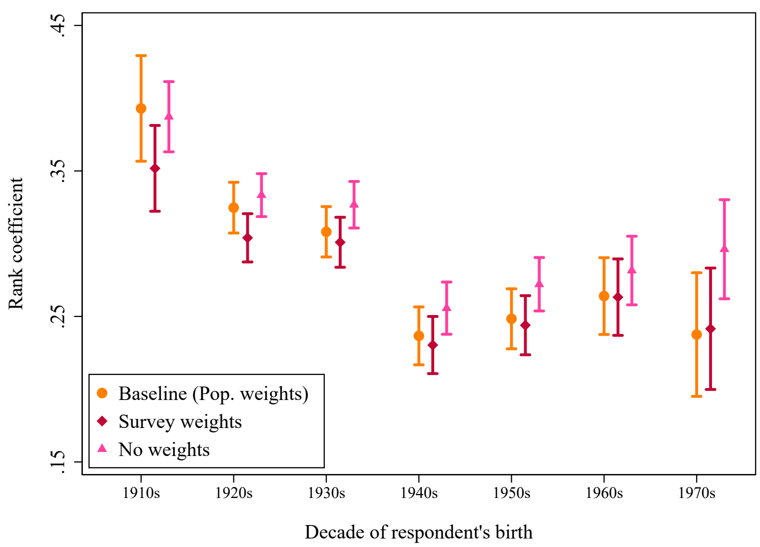
Notes: The first series in both panels reproduces the main IGE and rank-rank estimates using our baseline sample and income scores. In the second series, we bring in additional respondents from the 15 surveys whose father was present but not working (e.g., unemployed, retired). In the third series, we instead bring in additional respondents from the 15 surveys who provided information about their mother's occupation. The construction of income scores for non-working fathers and working mothers is detailed in Appendix D. The fourth series includes respondents who provided information about their non-working father *or* about their mother's occupation (and if both pieces of information were provided, we use income scores based on the mother's occupation). In the final series of the bottom panel, we assign all U.S.-born respondents ages 30–50 in our 15 surveys that still have missing parental income the lowest possible rank (i.e., assuming that their household had zero income in childhood).

Figure A.8: Mobility by birth decade, robustness to weights

(a) Intergenerational elasticity



(b) Rank-rank coefficient

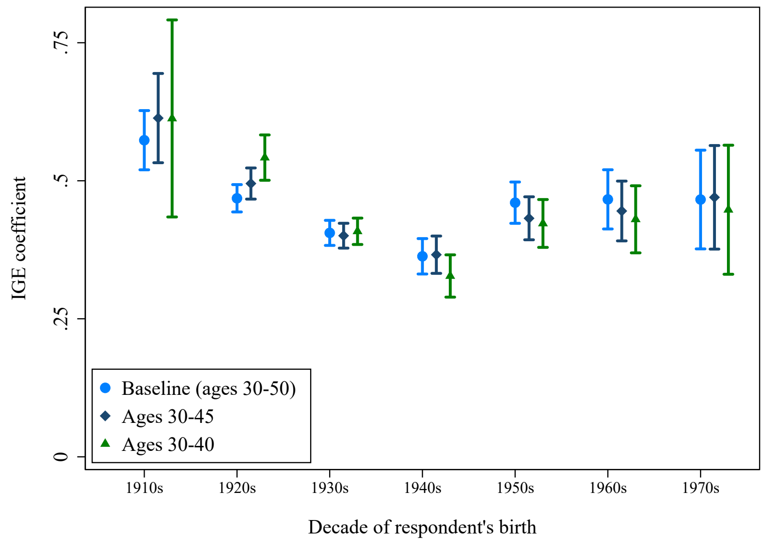


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

Notes: The estimates are based on the baseline sample of respondents ages 30–50. The first series in both panels reproduces the main IGE and rank-rank estimates using the baseline population-adjusted weights. In other words, in the first series, we re-weight survey weights so that each birth cohort has representative *race* $\times$ *sex* shares. The second series uses the provided survey weights (or a weight of one when no survey weight is available). The estimates from the third series are unweighted.

Figure A.9: Mobility by birth decade, robustness to age group

(a) Intergenerational elasticity



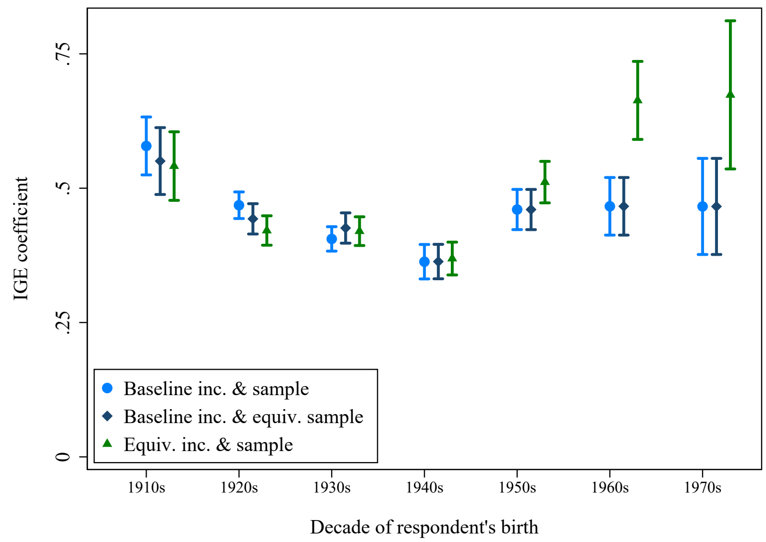
(b) Rank-rank coefficient

Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

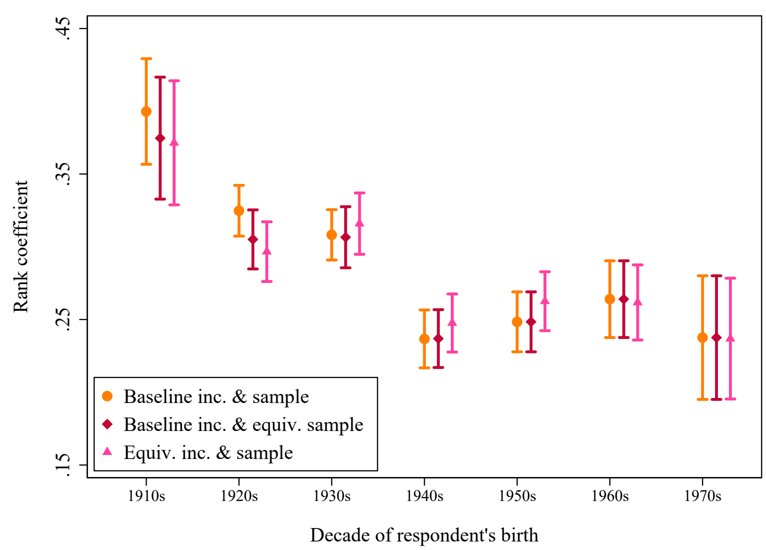
Notes: The estimates are based on the baseline sample of respondents aged 30–50. The first series in both panels reproduces the main IGE and rank-rank estimates using the baseline population-adjusted weights. In the second series, we restrict the sample to respondents aged 30–45. The third series further restricts the sample to those aged 30–40.

Figure A.10: Mobility by birth decade, robustness to family size

(a) Intergenerational elasticity



(b) Rank-rank coefficient

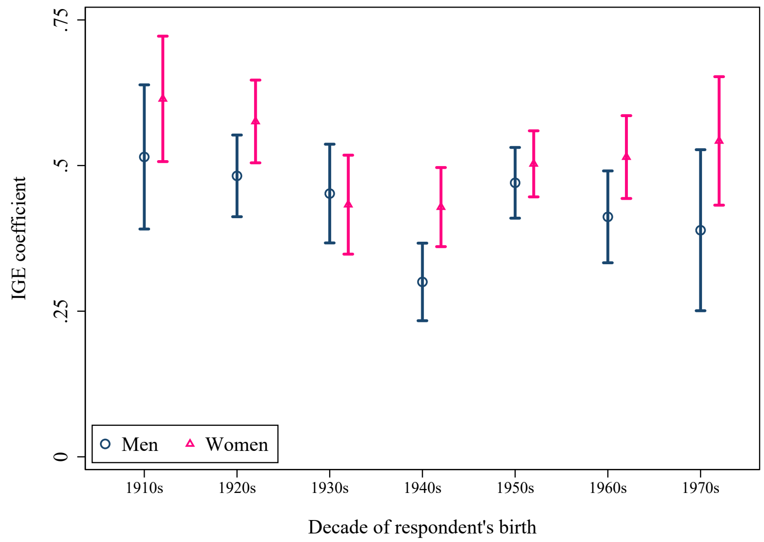


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D. We utilize data from Ruggles *et al.* (2021) to construct income scores and measures of household size.

Notes: The first series in both panels reproduces the main IGE and rank-rank estimates using our baseline sample and income scores. In the second series, we restrict the sample to respondents that provided information about their household size (which we refer to as the equivalized sample and is 84% of the baseline sample). In the third series, we use this sub-sample and adjust the income measures to account for differences in household size. For the respondent's generation, we divide own family income by the square root of a respondent's household size at the time of the interview. For the fathers' generation, we use the interpolated measure of a father's income and divide this income score by the square root of the median household size. Similarly to the interpolated measure of income, we use the 1920–1990 Censuses to construct the median household size when the respondent is 10 years old (taking the weighted average of the median household size in that *occupation* × *race* × *South* cell and allowing the weights to reflect the year in which the respondent is 10).

Figure A.11: Mobility measures by birth decade, by sex (restricted to common surveys)

(a) Intergenerational elasticity



(b) Rank-rank coefficient

Sources: This figure combines the 7 surveys that include both male and female respondents.

Notes: This figure is identical to Figure 7 except that in this figure, we only use surveys that include both men and women. In the second panel, respondents are ranked relative to other individuals in this sub-sample who are born in the same cohort. Fathers are ranked relative to all fathers in this sub-sample with a child in the same birth cohort. We use sample weights where provided and further weight each birth cohort in this sub-sample so that they have representative *race* × *sex* shares.

Table A.1: Summary Statistics in Panel Study of Income Dynamics

	1968 Men	1968 Fathers	With Child in Survey	Father's Income: 1 year	Father's Income: 5 years	Father's Income: 10 years
Age	40.06	39.98	39.75	39.60	38.36	36.26
Black	0.09	0.08	0.07	0.07	0.06	0.06
HS educated	0.56	0.58	0.61	0.62	0.62	0.65
College edu.	0.16	0.17	0.19	0.19	0.19	0.20
Family income	10,986	11,109	11,363	11,406	11,342	11,254
Observations	1,765	1,472	1,077	939	783	545

Notes: This table uses the Panel Study of Income Dynamics dataset from 1968 through 2015. The first column considers all men ages 30–50 in the 1968 wave of the PSID. Column 2 restricts that sample to household heads with children present in the family unit, away from home, or in an institution. Column 3 further restricts the sample to those who were identified by the PSID as the biological or adoptive fathers of other survey respondents using the Family Identification Mapping System (FIMS). Columns 4–6 then restrict the sample to fathers with 1, 5, and 10 years of available income between the ages of 30–50, respectively, and whose children had at least one year of available income between ages 30–50.

Table A.2: Differences in the 1910–1940 decline when looking only at white men versus representative samples

	IGE	Rank-rank
Difference between 1910s and 1940s cohorts:		
White men	-0.192	-0.121
	(0.040)	(0.020)
Representative sample	-0.213	-0.156
	(0.020)	(0.020)
P-value	0.553	0.086
Linear decline using 1910s–1940s cohorts:		
White men	-0.0060	-0.0039
	(0.0010)	(0.0006)
Representative sample	-0.0069	-0.0049
	(0.0006)	(0.0004)
P-value	0.370	0.081

Notes: The top panel considers the difference between respondents born in the 1910s birth cohorts and those born in the 1940s cohorts, using specifications like equations (1) and (2), but allowing the slope and intercept to differ by cohort. The reported coefficient and standard error correspond to the interaction term, which measures the difference in the slope between the two cohorts. In the second panel, we consider all respondents born in the 1910s–1940s cohorts and model the decline in the slope linearly. Specifically, we run specifications in which we interact father income (or rank) with a variable that measures the number of years between a respondent's birth and 1911 (including birth-year fixed effects). In both panels, the p-value corresponds to a test of whether the two coefficients (using white men versus representative samples) are equal using seemingly unrelated regressions.

B Alternative measures of father's income

In this Appendix section, we present alternative estimates of intergenerational mobility over the 20th century using different ways of measuring a father's income. In Figure B.1 we present results using the sub-sample of respondents that provide information on their father's education, using both our baseline income scores (which vary at the *occupation* $\times$ *race* $\times$ *South* level) and income scores that incorporate a father's educational attainment.[29] The IGE and rank-rank estimates are very similar to our baseline case, with the IGE estimates slightly lower during the first half of the 20th century and slightly higher than the baseline in the last two cohorts, while the rank-rank is slightly higher throughout the sample period.

In Figure B.2, we again focus on the sub-sample of respondents that provide information about their father's educational attainment, showing how the IGE and rank-rank estimates change as we add more information about the father in the income-score construction. The first series only allows the scores to vary by occupation, but the second, third, and fourth series successively incorporate detail on the race, region, and educational attainment of the father. We see that the decline in the IGE and rank-rank estimates between the 1910s and 1940s cohorts is remarkably unchanged despite adding important predictors to the income scores.[30] Table B.1 summarizes the results from this exercise, quantifying the decline between the 1910s and 1940s birth cohorts using these alternative ways of measuring a father's income.

In Figure B.3 we present results that adjust the income of fathers who were farmers. We adjust this income score in two separate ways relative to the baseline measure. First, we use the 1900 Census of Agriculture to estimate the income of fathers who are farmers, and we allow this measure to vary at the *race* $\times$ *South* level and by the share of individuals in that cell that are farm owners (for more details, we refer the reader to Appendix D). In our second adjustment, we simply exclude respondents whose fathers are farmers, and we re-rank the remaining respondents and their fathers. As Figure B.3 shows, the IGE pattern remains consistent with both of these adjustments, while the rank-rank correlations during the first half of the 20th century fall when farmers are excluded. Nonetheless the decline from 1910 to 1940 is still visible and statistically significant. Given that the United States is experiencing significant structural change out of agriculture in this period, it is remarkable that the exclusion of farmer fathers has such a small effect on the overall pattern of mobility.

Finally, in Figure B.4 we present results that use alternative data sources for or alternative ways of constructing fathers' income scores. Recall that our baseline income scores rely on measures of household income from the 1940 Census (reproduced in the first series). We alter this preferred measure in a number of ways (for more details

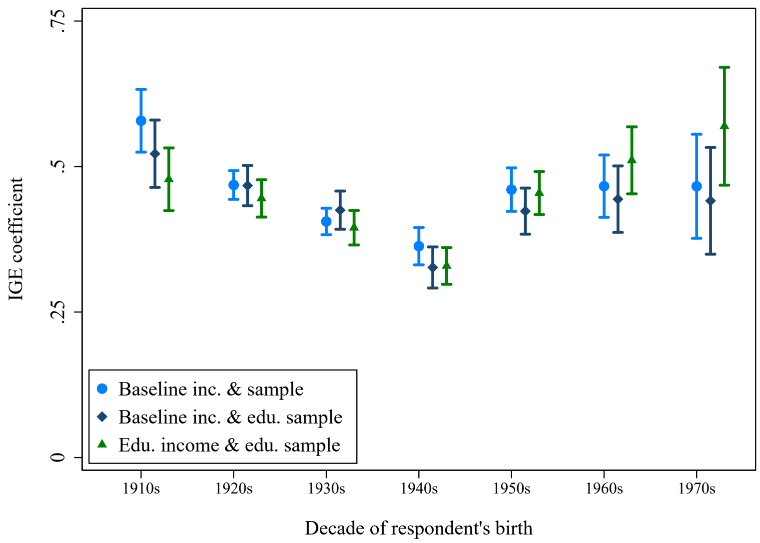
[29] Of the 15 surveys in our baseline sample, 12 include information about a father's educational attainment, representing nearly 80 percent of the baseline sample.

[30] An increase in the level of the rank-rank coefficient is unsurprising in this setting given that incorporating additional information into the income score likely increases the covariance between the rank of children and the rank of their fathers, while leaving the variance of the father's ranks relatively unchanged (by construction, given that ranks range from 0 to 100).

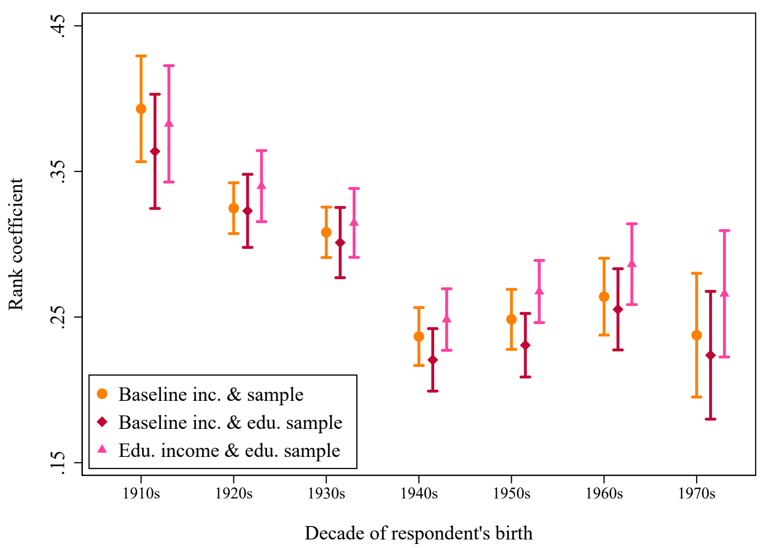
on the construction of all of these series, we refer the reader to Appendix D). First, instead of using measures of *household* income, we simply use individual-level wage information for fathers ages 30–50 to construct the income scores (second series). Next, we use income data from around 1900 to construct income scores, combining the 1901 Cost of Living Survey and the 1900 Census of Agriculture. Third, we construct an "interpolated" measure of father's income that uses the data source that is closest in time to the respondent's tenth birthday (e.g., for respondents born in 1921, we assign an income score that is a weighted average of the 1900- and 1940-based income scores, with the weights reflecting the number of years between 1931 and 1940). Finally, we compare our results to simply using the 1950 IPUMS *occscore* variable (which, recall, pooled all adults and computed the nation-wide median income for each occupation). The IGE measures are higher for the 1900 and the interpolated measures of father's income, but the overall u-shape remains stark and salient. The rank-rank measures all look quite comparable to each other, with the exception being the series that uses the 1950 IPUMS *occscore* variable, thereby highlighting the value of using occupational scores for fathers that incorporate race and region.

Figure B.1: Mobility by birth decade, adjusting father income score for education

(a) Intergenerational elasticity



(b) Rank-rank coefficient

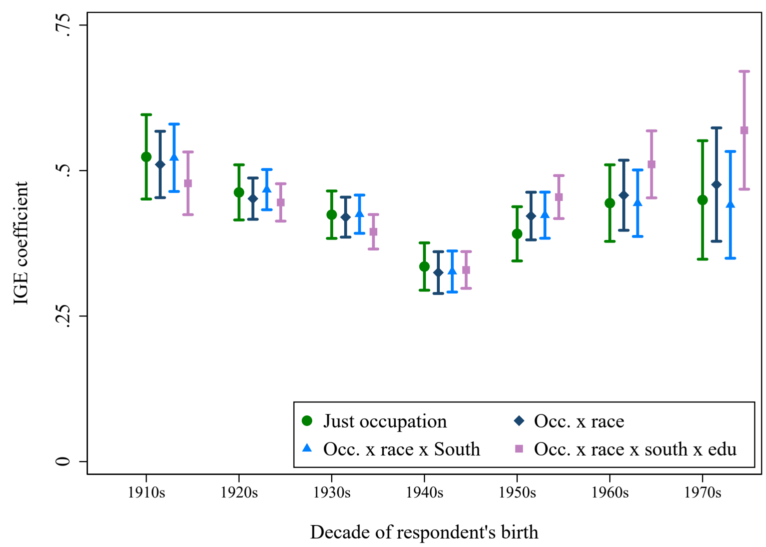


Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

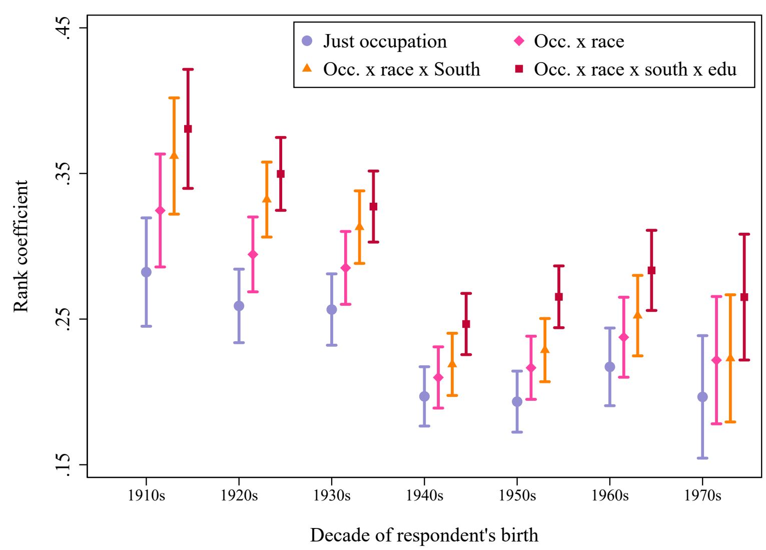
Notes: The first series in both panels reproduces the main IGE and rank-rank estimates using our baseline sample and income scores. In the second series, we continue to use the baseline income scores, but restrict the sample to respondents ages 30–50 who provided information on their fathers' education (available in 12 of the 15 surveys). In the third series, we use this smaller sub-sample in conjunction with income scores that also vary by a father's educational attainment. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) in all samples so that they have representative *race* × *sex* shares.

Figure B.2: Mobility measures by birth decade, adding detail to fathers' income scores

(a) Intergenerational elasticity



(b) Rank-rank coefficient

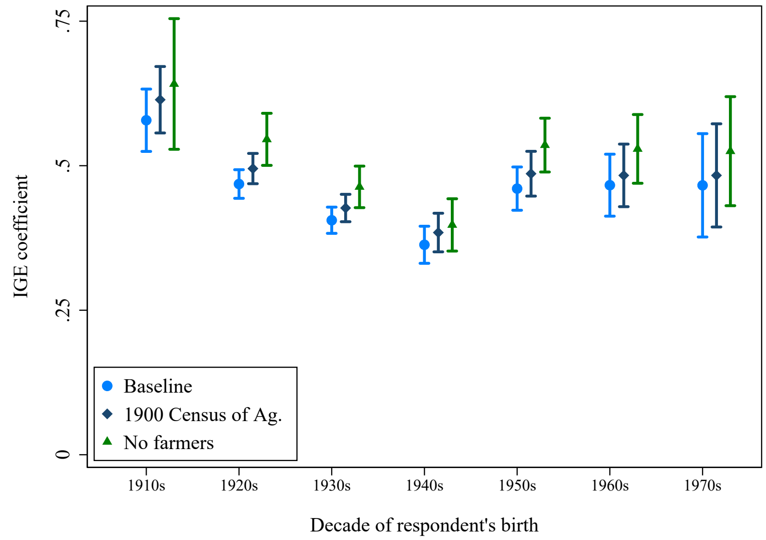


Sources: This figure combines 12 surveys in which respondents provide information on a father's educational attainment. Further detail is available in Appendix D.

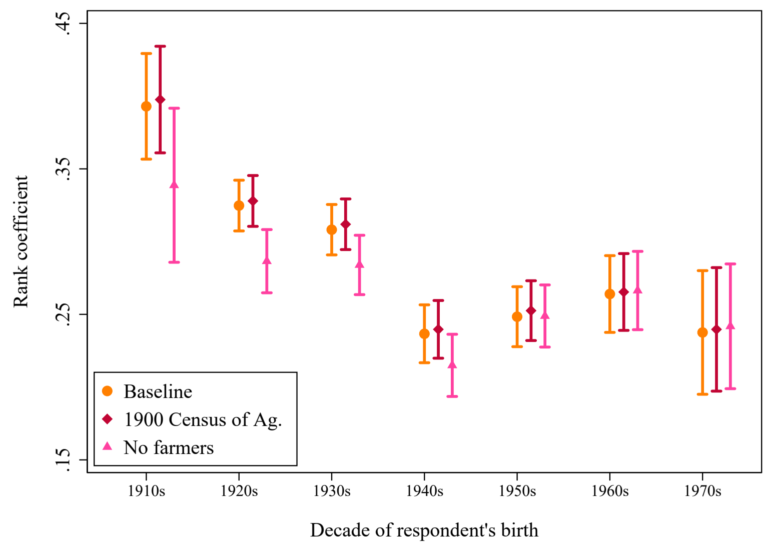
Notes: The IGE and rank-rank are based on the sample of respondents ages 30–50 who provided information about their fathers' educational attainment. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) in this sample so that they have representative *race* × *sex* shares. The first series uses income scores that only vary by a father's occupation. The second series allows income scores to vary by father's occupation and race. The third series allows income scores to vary by father's occupation, race, and Southern residence. The fourth series allows income scores to vary by father's occupation, race, Southern residence, and father's educational level.

Figure B.3: Mobility by birth decade, incorporating various adjustments for farmers

(a) Intergenerational elasticity



(b) Rank-rank coefficient



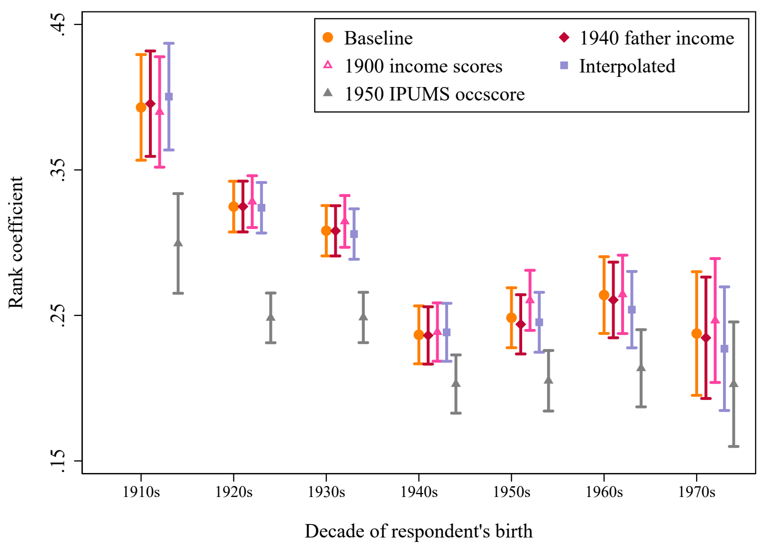
Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

Notes: The first series in both panels reproduces the main IGE and rank-rank estimates using our baseline income scores. The second series uses the 1900 Census of Agriculture to estimate income for fathers who are farmers. In both the first and second series, the IGE and rank-rank estimates are based on the baseline sample of respondents aged 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares. The third series drops all respondents whose fathers work in agricultural occupations; the remaining respondents are re-ranked in this sub-sample, and weights are constructed so that each birth cohort in this sub-sample also has representative *race* × *sex* shares.

Figure B.4: Mobility by birth decade, incorporating various income score adjustments

(a) Intergenerational elasticity

(b) Rank-rank coefficient



Sources: This figure combines 15 different surveys, which are described in Section 2 and in further detail in Appendix D.

Notes: All estimates in this figure are based on the baseline sample of respondents ages 30–50. We use sample weights where provided and further re-weight each birth cohort in this sample so that they have representative *race* × *sex* shares. The first series in both panels reproduces the main IGE and rank-rank estimates using our baseline income scores. "1940 father income" refers to using median individual (as opposed to household) income for fathers with a certain occupation, race, and Southern residence. "1900 income scores" refers to using the 1900 Census of Agriculture in conjunction with the 1901 Cost of Living Survey to construct income scores. "Interpolated" refers to using the dataset closest in time to when the respondent was 10 years old to predict father income. "1950 IPUMS occscore" refers to using the *occscore* variable from IPUMS. For more detail on the construction of these income scores, see Appendix D.

Table B.1: Quantifying the 1910–1940 decline in the IGE and rank coefficient as we add information about fathers to the income scores

(a) Intergenerational elasticity

	(1) Occupation	(2) Occ. x race	(3) Occ. x race x South	(4) Occ. x race x South x edu
Father income	0.5581***	0.5415***	0.5604***	0.5119***
	[0.0313]	[0.0243]	[0.0240]	[0.0220]
Inc. x (Year-1910)	-0.0064***	-0.0061***	-0.0065***	-0.0052***
	[0.0013]	[0.0010]	[0.0010]	[0.0009]
Observations	31,088	31,088	31,088	31,088

(b) Rank-rank coefficient

	(1) Occupation	(2) Occ. x race	(3) Occ. x race x South	(4) Occ. x race x South x edu
Father rank	0.3083***	0.3598***	0.4125***	0.4205***
	[0.0162]	[0.0166]	[0.0168]	[0.0165]
Rank x (Year-1910)	-0.0030***	-0.0041***	-0.0053***	-0.0048***
	[0.0006]	[0.0007]	[0.0007]	[0.0007]
Observations	31,088	31,088	31,088	31,088

Notes: All estimates are based on the sample of respondents ages 30–50 who provided information about their fathers' educational attainment and were born before 1950. Each column varies the information used to construct income scores for fathers. In all specifications, we interact father income (or rank) with a variable that measures the number of years between a respondent's birth year and 1911. All specifications include survey-year and birth-year fixed effects. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* $\times$ *sex* shares.

C Assessing recall bias

Our estimates of mobility rely on survey respondents' recollection of their fathers' occupations. In this section, we consider the extent to which recall bias might be present in our estimates. We begin by comparing the fathers' occupations provided by male versus female respondents in our surveys. We then compare the fathers' occupations in our surveys to those of fathers in the decennial Censuses at the time that the respondents were growing up. We conclude by looking at the PSID—a survey that includes both retrospective questions as well as self-reported information about fathers' occupations when the respondent was growing up—to gauge the extent to which adult children's retrospective answers match fathers' self-reported occupations.

C.1 Comparing male and female survey respondents

We begin by comparing the fathers' occupations reported by male and female respondents in our surveys. Roughly speaking, brothers and sisters grow up in the same families in the US, so adult men and women should report similar occupations for their fathers. Strictly speaking, small differences could arise between the average income score of men and women. If parents have sex-based stopping rules when making fertility decisions, then boys and girls might wind up growing up in systematically different families (as in Asher *et al.* 2018, using data from India). However, evidence for sex-based fertility patterns in the US is much weaker. Second, even if boys and girls grow up in identical families in terms of parental income, small differences might arise because men have higher mortality rates than women and thus selection into surviving into prime age could differ by gender (especially in our oldest cohorts, men are less likely to live until age 50).

These small potential discrepancies notwithstanding, we would be suspicious of any parental income estimate that gives significantly different estimates for male and female respondents. We thus regress the log as well as the rank of estimated parental income on a female dummy, separately for each of our birth decades, and report the results in Appendix Table C.1. The coefficient on the female dummy is always close to zero and has no consistent sign. We repeat this analysis separately for white and Black respondents and report the results in Appendix Table C.2. Again, we find no notable patterns or significant differences beyond what might be expected by chance. Appendix Table C.3 shows the top five occupations reported by male and female respondents in each birth cohort. In all birth cohorts, at least four—if not all five—of the top occupations coincide between male and female respondents and in roughly the same order.

Table C.1: Differences in Income Scores, by Respondent Sex and Birth Cohort

(a) Logged Father's Income

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
Female	0.003	0.032	-0.019	0.007	-0.003	-0.013	-0.001
	[0.035]	[0.022]	[0.020]	[0.013]	[0.010]	[0.011]	[0.018]
Observations	5,207	13,328	12,446	11,575	10,962	6,598	3,125

(b) Ranked Father's Income

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
Female	-0.467	1.955	-1.711	0.536	-0.208	-1.012	-0.457
	[1.846]	[1.271]	[1.136]	[0.786]	[0.669]	[0.790]	[1.245]
Observations	5,207	13,328	12,446	11,575	10,962	6,598	3,125

Notes: This table uses our baseline sample ages 30–50 to regress logged and ranked father's income on an indicator variable for whether a respondent is female. Age and age squared controls are included in the first panel, and survey-year fixed effects are included in both panels. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

Table C.2: Differences in Income Scores, by Respondent Sex, Race and Birth Cohort

(a) Logged Father's Income, White Respondents

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
Female	-0.002 [0.027]	0.035* [0.019]	-0.024 [0.019]	-0.007 [0.012]	0.002 [0.009]	-0.014 [0.010]	-0.014 [0.016]
Observations	4,605	11,653	10,637	9,837	9,034	5,518	2,389

(b) Ranked Father's Income, White Respondents

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
Female	-0.376 [1.800]	1.984 [1.299]	-2.111* [1.230]	-0.099 [0.828]	-0.179 [0.673]	-1.167 [0.795]	-1.037 [1.259]
Observations	4,605	11,653	10,637	9,837	9,034	5,518	2,389

(c) Logged Father's Income, Black Respondents

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
Female	0.095 [0.145]	-0.084 [0.072]	0.017 [0.051]	0.022 [0.036]	-0.039 [0.026]	-0.017 [0.030]	0.098** [0.040]
Observations	602	1,675	1,809	1,738	1,928	1,080	736

(d) Ranked Father's Income, Black Respondents

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
Female	0.908 [4.653]	-2.294 [2.376]	1.095 [1.520]	1.092 [1.000]	-0.111 [0.847]	-0.215 [1.098]	4.923*** [1.586]
Observations	602	1,675	1,809	1,738	1,928	1,080	736

Notes: This table uses our baseline sample of ages 30–50 to regress logged and ranked father's income on an indicator variable for whether a respondent is female, separately by respondent race. Age and age squared controls are included in the first panel, and survey-year fixed effects are included in both panels. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

Table C.3: Top Five Occupations Reported by Male and Female Respondents, by Birth Cohort

Birth Cohort	Male Respondents	Share of male sample	Female Respondents	Share of female sample
1910s	1. Farm operator	0.33	1. Farm operator	0.36
	2. Craftsman (skilled)	0.15	2. Craftsman (skilled)	0.16
	3. Craftsman (semi-skilled)	0.15	3. Craftsman (semi-skilled)	0.12
	4. Unskilled laborer (non-farm)	0.07	4. Unskilled laborer (non-farm)	0.08
	5. Businessman (self-employed)	0.05	5. Businessman (self-employed)	0.07
1920s	1. Farm operator	0.25	1. Farm operator	0.25
	2. Craftsman (skilled)	0.18	2. Craftsman (semi-skilled)	0.19
	3. Craftsman (semi-skilled)	0.16	3. Craftsman (skilled)	0.14
	4. Unskilled laborer (non-farm)	0.08	4. Businessman (not self-employed)	0.10
	5. Businessman (self-employed)	0.06	5. Unskilled laborer (non-farm)	0.07
1930s	1. Farm operator	0.19	1. Farm operator	0.19
	2. Craftsman (skilled)	0.18	2. Craftsman (semi-skilled)	0.19
	3. Craftsman (semi-skilled)	0.18	3. Craftsman (skilled)	0.17
	4. Unskilled laborer (non-farm)	0.08	4. Businessman (not self-employed)	0.11
	5. Businessman (self-employed)	0.06	5. Unskilled laborer (non-farm)	0.07
1940s	1. Craftsman (skilled)	0.21	1. Craftsman (semi-skilled)	0.19
	2. Craftsman (semi-skilled)	0.17	2. Craftsman (skilled)	0.19
	3. Farm operator	0.12	3. Businessman (not self-employed)	0.11
	4. Businessman (not self-employed)	0.11	4. Farm operator	0.10
	5. Unskilled laborer (non-farm)	0.06	5. Unskilled laborer (non-farm)	0.07
1950s	1. Craftsman (skilled)	0.20	1. Craftsman (skilled)	0.19
	2. Craftsman (semi-skilled)	0.16	2. Craftsman (semi-skilled)	0.17
	3. Businessman (not self-employed)	0.14	3. Businessman (not self-employed)	0.12
	4. Farm operator	0.06	4. Unskilled laborer (non-farm)	0.06
	5. Unskilled laborer (non-farm)	0.06	5. Farm operator	0.06
1960s	1. Craftsman (skilled)	0.20	1. Craftsman (skilled)	0.20
	2. Craftsman (semi-skilled)	0.16	2. Craftsman (semi-skilled)	0.17
	3. Businessman (not self-employed)	0.14	3. Businessman (not self-employed)	0.13
	4. Unskilled laborer (non-farm)	0.05	4. Unskilled laborer (non-farm)	0.05
	5. Protective service officer	0.05	5. Protective service officer	0.05
1970s	1. Craftsman (skilled)	0.18	1. Craftsman (semi-skilled)	0.17
	2. Craftsman (semi-skilled)	0.14	2. Craftsman (skilled)	0.18
	3. Businessman (not self-employed)	0.13	3. Businessman (not self-employed)	0.14
	4. Protective service officer	0.08	4. Unskilled laborer (non-farm)	0.05
	5. Unskilled laborer (non-farm)	0.06	5. Protective service officer	0.06

Notes: Estimates are based on the baseline sample of respondents ages 30–50. We use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares.

C.2 Comparing survey respondents' fathers to Census fathers

Next, we compare the occupations of fathers in the surveys to those of actual fathers in the Census in the years when the respondents were children. Ward (2020) warns that Census-takers made errors in recording the occupation variable, but we would still be worried if our respondents' recollection of their fathers' occupations differed dramatically from fathers' occupations in the Census during the years in which the respondents were growing up. In all of the exercises in this section, we consider both the earlier and later corresponding Censuses, when respondents were ages 0–10 and 11–20, respectively.

We begin by comparing the average income scores of fathers in the surveys with the average income scores of fathers in the Census. Appendix Table C.4 regresses the estimates of logged parental income on a dummy for whether the father's income score came from the surveys.[31] We note that the average income scores of fathers in the surveys are consistently lower than those of fathers in the Census, but the point estimates are small. More importantly, there does not seem to be any pattern in how the estimates are changing, suggesting that recall bias is not improving or deteriorating across cohorts. This lack of a consistent pattern, especially in the first half of cohorts, suggests that the rise in mobility that we find is not driven by respondents' remembering their fathers' occupations differently across cohorts (or in other words, it does not seem to be the case that the rise in mobility is driven by measurement error changing monotonically over time).

Appendix Tables C.5 and C.6 compare the mix of coarsened occupations that our respondents report their fathers as having to that of fathers in the Census. In both of these tables, we find that the share of fathers with each occupation are comparable to the corresponding shares in the Census.[32]

[31] We do not include an analogous exercise using ranked father's income. When calculating ranks for fathers in our main analysis, we rank a survey respondent's father relative to all fathers with children born in the same birth cohort. Because we are comparing these men to fathers in the decennial Census (most of whom have multiple children), it is not obvious which child's year of birth should be used in the ranking. Similarly, because we do not know the exact age of survey respondents' fathers, we cannot rank survey and Census fathers using their age.

[32] The share of survey fathers who are farmers exceeds the analogous Census share in Appendix Table C.6. However, it is worth noting that the decline of agriculture as a dominant occupation was occurring during this time period, so we would expect the Census shares to be lower than the survey shares when considering the later Census.

Table C.4: Differences in Logged Income Scores between Census Fathers and Survey Respondents' Fathers

(a) Using Earlier Census

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960	(7) 1970
Survey respondent	-0.034***	-0.060***	-0.034***	-0.018***	-0.008*	-0.026***	-0.031***
	[0.012]	[0.006]	[0.005]	[0.005]	[0.005]	[0.006]	[0.009]
Observations	93,327	116,925	119,052	194,098	170,372	170,061	172,637

(b) Using Later Census

	(1) 1910	(2) 1920	(3) 1930	(4) 1940	(5) 1950	(6) 1960
Survey respondent	-0.102***	-0.090***	-0.132***	-0.095***	-0.055***	-0.041***
	[0.012]	[0.006]	[0.005]	[0.005]	[0.005]	[0.006]
Observations	108,804	119,934	194,969	170,985	174,425	176,110

Notes: Column titles reflect the decade of birth for survey respondents. Survey respondents include the baseline sample of respondents ages 30–50. Census respondents are restricted to fathers ages 30–50 in a particular Census year. In the top panel, the survey respondents' fathers are compared to the fathers in the Census when the respondents would have been between the ages of 1 and 10 (e.g., the fathers of survey respondents born in 1910–1919 are compared to 1920 Census fathers). In the bottom panel, the survey respondents' fathers are compared to fathers in the Census when the respondents would have been between the ages of 11 and 20 (e.g., the fathers of survey respondents born in 1910–1919 are compared to 1930 Census fathers).

Table C.5: Occupations of Survey Respondents' Fathers and Census Fathers (Using Earlier Census), by Birth Cohort

	1910–1919		1920–1929		1930–1939		1940–1949		1950–1959		1960–1969	
	Census (1920)	Survey	Census (1930)	Survey	Census (1940)	Survey	Census (1950)	Survey	Census (1960)	Survey	Census (1970)	Survey
Coarsened Occupations												
Accountants and auditors	0.34	0.53	0.52	0.47	0.62	0.57	0.88	0.62	0.99	1.16	1.11	0.99
Clergymen	0.46	0.67	0.41	0.73	0.41	0.63	0.40	0.60	0.44	0.77	0.54	0.56
Public-school teachers	0.34	0.64	0.48	0.46	0.83	0.52	0.85	0.88	1.14	1.38	2.06	2.15
Dentists	0.19	0.29	0.24	0.16	0.27	0.24	0.21	0.13	0.21	0.24	0.23	0.19
Physicians and surgeons	0.51	0.67	0.40	0.27	0.49	0.42	0.59	0.48	0.62	0.90	0.70	0.56
Engineers	0.53	1.08	0.72	0.87	0.70	1.07	1.56	2.10	2.58	3.41	3.61	3.90
Lawyers and judges	0.44	0.31	0.45	0.42	0.56	0.48	0.51	0.48	0.51	0.64	0.67	0.74
Social and welfare workers	0.04	0.02	0.03	0.03	0.04	0.08	0.08	0.08	0.12	0.12	0.19	0.21
Nurses (trained or student)	0.00	0.01	0.01	0.05	0.02	0.04	0.01	0.08	0.03	0.06	0.15	0.04
Other professional and technical	0.58	0.40	0.74	0.66	1.02	0.96	1.61	1.76	2.43	3.34	4.61	3.91
Semi-professional	0.69	0.84	0.88	0.66	0.91	1.17	1.49	1.72	2.35	2.21	3.08	2.69
Businessmen (self-employed)	6.44	6.26	6.35	4.25	4.73	3.84	6.52	3.17	4.29	2.82	3.30	3.00
Businessmen (not self-employed)	3.90	4.65	5.24	7.21	5.50	7.87	6.18	11.33	8.09	12.97	9.39	13.68
Bookkeeper	0.48	0.19	0.38	0.34	0.41	0.32	0.30	0.26	0.25	0.19	0.44	0.16
Stenographers	0.08	0.27	0.14	0.18	0.12	0.09	0.16	0.14	0.12	0.11	0.17	0.05
Other clerical workers	3.17	1.64	3.41	2.88	4.27	2.95	4.83	3.67	5.28	3.53	5.01	3.08
Sales: higher-status	0.96	1.30	1.41	1.07	1.01	0.99	1.11	1.24	1.52	1.73	2.08	2.13
Sales: inside sales	2.93	1.88	4.33	2.16	6.96	2.64	4.85	3.54	5.09	3.72	4.81	3.89
Sales: lower-status	0.17	0.38	0.19	0.18	0.13	0.20	0.05	0.07	0.05	0.06	0.08	0.05
Foremen	1.96	1.74	2.14	2.19	1.89	3.11	2.62	3.27	3.30	3.77	3.92	3.77
Craftsmen (skilled)	17.10	15.57	17.17	16.20	15.58	17.98	18.16	19.72	19.03	19.38	19.01	19.93
Craftsmen (semi-skilled)	13.46	13.44	15.07	17.68	17.65	18.36	20.41	18.08	20.46	16.68	18.90	16.87
Protective service officers	0.96	1.18	1.32	2.07	1.45	2.16	2.35	3.53	3.72	4.49	4.19	4.85
Private household workers	0.10	0.03	0.09	0.66	0.25	0.36	0.04	0.25	0.03	0.07	0.03	—
Other service workers	1.84	1.90	2.44	2.73	2.96	2.92	2.54	2.82	2.41	2.82	3.25	2.55
Farm laborers	3.20	1.92	3.37	2.94	4.24	3.45	2.13	2.73	1.37	1.67	1.03	1.13
Unskilled non-farm laborers	10.71	7.65	10.95	7.67	12.19	7.40	6.27	6.42	5.47	5.69	4.47	5.25
Farm operators	27.22	34.54	20.18	24.80	14.79	19.19	10.73	10.82	5.01	6.07	2.55	3.65

Notes: For survey estimates, we use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares. Census shares are weighted using provided weights. Census samples include all Black and white fathers ages 30–50. The fathers in the Census are compared to survey respondents' fathers when the respondents would have been between the ages of 1 and 10 at the time of the Census (e.g., survey respondents born in 1910–1919 are compared to 1920 Census respondents.)

Table C.6: Occupations of Survey Respondents' Fathers and Census Fathers (Using Later Census), by Birth Cohort

	1910–1919		1920–1929		1930–1939		1940–1949		1950–1959		1960–1969	
	Census (1930)	Survey	Census (1940)	Survey	Census (1950)	Survey	Census (1960)	Survey	Census (1970)	Survey	Census (1980)	Survey
Coarsened Occupations												
Accountants and auditors	0.52	0.53	0.62	0.47	0.88	0.57	0.99	0.62	1.11	1.16	1.05	0.99
Clergymen	0.41	0.67	0.41	0.73	0.40	0.63	0.44	0.60	0.54	0.77	0.57	0.56
Public-school teachers	0.48	0.64	0.83	0.46	0.85	0.52	1.14	0.88	2.06	1.38	2.83	2.15
Dentists	0.24	0.29	0.27	0.16	0.21	0.24	0.21	0.13	0.23	0.24	0.28	0.19
Physicians and surgeons	0.40	0.67	0.49	0.27	0.59	0.42	0.62	0.48	0.70	0.90	0.79	0.56
Engineers	0.72	1.08	0.70	0.87	1.56	1.07	2.58	2.10	3.61	3.41	2.96	3.90
Lawyers and judges	0.45	0.31	0.56	0.42	0.51	0.48	0.51	0.48	0.67	0.64	0.92	0.74
Social and welfare workers	0.03	0.02	0.04	0.03	0.08	0.08	0.12	0.08	0.19	0.12	0.29	0.21
Nurses (trained or student)	0.01	0.01	0.02	0.05	0.01	0.04	0.03	0.08	0.15	0.06	0.24	0.04
Other professional and technical	0.74	0.40	1.02	0.66	1.61	0.96	2.43	1.76	4.61	3.34	4.39	3.91
Semi-professional	0.88	0.84	0.91	0.66	1.49	1.17	2.35	1.72	3.08	2.21	3.55	2.69
Businessmen (self-employed)	6.35	6.26	4.73	4.25	6.52	3.84	4.29	3.17	3.30	2.82	3.73	3.00
Businessmen (not self-employed)	5.24	4.65	5.50	7.21	6.18	7.87	8.09	11.33	9.39	12.97	12.31	13.68
Bookkeeper	0.38	0.19	0.41	0.34	0.30	0.32	0.25	0.26	0.44	0.19	0.20	0.16
Stenographers	0.14	0.27	0.12	0.18	0.16	0.09	0.12	0.14	0.17	0.11	0.08	0.05
Other clerical workers	3.41	1.64	4.27	2.88	4.83	2.95	5.28	3.67	5.01	3.53	5.08	3.08
Sales: higher-status	1.41	1.30	1.01	1.07	1.11	0.99	1.52	1.24	2.08	1.73	2.01	2.13
Sales: inside sales	4.33	1.88	6.96	2.16	4.85	2.64	5.09	3.54	4.81	3.72	4.06	3.89
Sales: lower-status	0.19	0.38	0.13	0.18	0.05	0.20	0.05	0.07	0.08	0.06	0.10	0.05
Foremen	2.14	1.74	1.89	2.19	2.62	3.11	3.30	3.27	3.92	3.77	4.55	3.77
Craftsmen (skilled)	17.17	15.57	15.58	16.20	18.16	17.98	19.03	19.72	19.01	19.38	17.13	19.93
Craftsmen (semi-skilled)	15.07	13.44	17.65	17.68	20.41	18.36	20.46	18.08	18.90	16.68	16.97	16.87
Protective service officers	1.32	1.18	1.45	2.07	2.35	2.16	3.72	3.53	4.19	4.49	4.46	4.85
Private household workers	0.09	0.03	0.25	0.66	0.04	0.36	0.03	0.25	0.03	0.07	0.01	—
Other service workers	2.44	1.90	2.96	2.73	2.54	2.92	2.41	2.82	3.25	2.82	3.08	2.55
Farm laborers	3.37	1.92	4.24	2.94	2.13	3.45	1.37	2.73	1.03	1.67	0.79	1.13
Unskilled non-farm laborers	10.95	7.65	12.19	7.67	6.27	7.40	5.47	6.42	4.47	5.69	4.44	5.25
Farm operators	20.18	34.54	14.79	24.80	10.73	19.19	5.01	10.82	2.55	6.07	1.97	3.65

Notes: For survey estimates, we use sample weights where provided and further re-weight each birth cohort (i.e., decade) so that they have representative *race* × *sex* shares. Census shares are weighted using provided weights. Census samples include all Black and white fathers ages 30–50. The fathers in the Census are compared to survey respondents' fathers when the respondents would have been between the ages of 11 and 20 at the time of the Census (e.g., survey respondents born in 1910–1919 are compared to 1930 Census respondents.)

C.3 Assessing recall bias using the PSID

In this section, we utilize the unique nature of the PSID to consider the extent to which retrospective questions convey accurate information about a father's income level. In particular, the PSID has both retrospective questions about a respondent's father's occupation as well as self-reported father's occupations and income when the respondent was growing up (i.e., in earlier waves of the survey). We focus on the 1997–2015 waves of the PSID because 1997 is the first year in which the retrospective questions are asked with sufficient detail (i.e., 3-digit occupation codes), so that they can be mapped to our coarsened occupations.

The way that we verify the retrospective answers is by looking at individuals who were household heads at some point between 1997–2015 and who were thus asked about their father's occupation while they were growing up. Then, using the Family Identification Mapping System (FIMS) provided by the PSID, we can find these individuals' fathers in earlier waves of the survey and see the fathers' self-reported (coarsened) occupations between the ages of 25–50 (i.e., when the respondents were growing up). We can then see whether the retrospective answers in 1997–2015 matched any of the self-reported occupations in earlier survey years.

We find that for 81% of adult children, their retrospective answers coincided with one of the self-reported occupations of their fathers during their childhood.[33] We can also then see what the most common mistakes were in identifying occupation (in other words, conditional on a respondent mis-reporting his/her father's occupation, what did the adult child typically report versus what did the father typically report). The four most common mistakes—which account for roughly 20% of all mistakes—are the respondents reporting that their fathers were skilled craftsmen, semi-skilled craftsmen, or unskilled non-farm laborers, when instead the father reported one of the other occupations on this same list.[34]

Even if one-in-five respondents are mis-reporting their fathers' occupations, it might still be the case that the retrospective answers convey accurate information about a father's income level. Appendix Figure C.1 plots the income score of fathers using the retrospective answers against the income score of fathers using self-reported occupations when they were around 40 years old. Both panels of this figure confirm that respondents' retrospective answers are highly correlated with fathers' self-reported answers, and thus convey similar information about the respondents' income level during their upbringing. Appendix Table C.7 regresses the five-year average of a father's self-reported actual income on alternative ways of measuring that father's income level. The coefficient of 1 in column 2, which uses the retrospective answer provided by the adult child, highlights that the retrospective answers seem to be reliable measures of a

[33] There are some instances (roughly 10% of respondents) in which the adult children's retrospective answers change across waves (for example, as a result of re-interviews due to changing family composition), so we consider all of the retrospective answers provided.

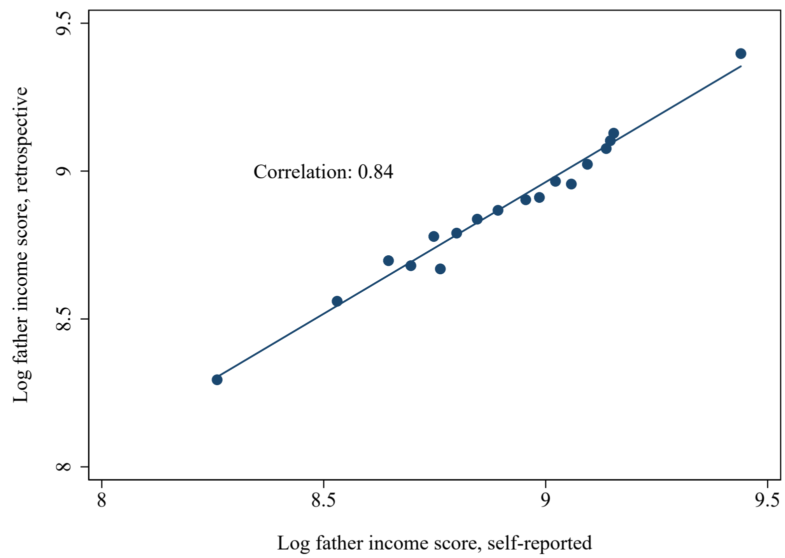
[34] To find the most common mistakes, we first select the 20% of respondents who were unable to accurately report any of their fathers' occupations. We then compare the modal retrospective answer in the data to the modal self-reported occupation of fathers between the ages of 30–50.

father's permanent income.

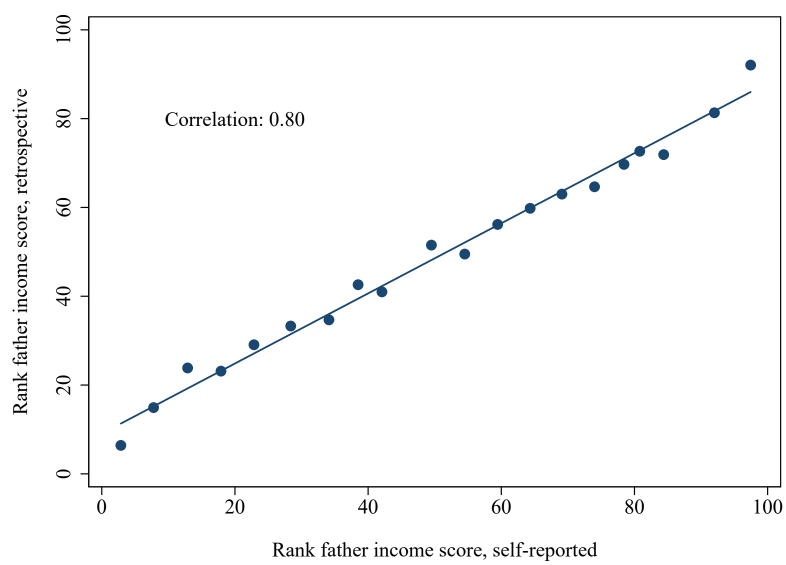
Finally, most of the estimates in the historical intergenerational mobility literature come from linked Census data (20 or 30 years apart) and use one year of a father's occupation to proxy for that father's income. To get a better sense of how estimates that use our retrospective approach differ from those that use the typical approach in the literature, Appendix Table C.8 estimates the IGE and rank-rank coefficient based on adult children's retrospective answers in 1997 as well as one year of father's self-reported occupation roughly thirty years earlier (around 1970). The mobility estimates are remarkably similar using the two approaches. Ward (2020) notes the measurement error that can arise from using one year of father's occupation, so the last column of this table also reports the mobility estimates from averaging the income scores of three self-reported father's occupations around 1970.

Figure C.1: Comparing fathers' income scores using adult children's retrospective answers and father's self-reported occupations

(a) Logged income scores



(b) Rank income scores



Sources: This figure uses the 1968–2015 Panel Study of Income Dynamics.

Notes: These figures are bin-scatter depictions of the income scores of fathers using the retrospective answers (y-axis) versus the income scores using fathers' self-reported answers in earlier survey waves (x-axis). The y-axis assigns fathers the 1970-based income score using the modal retrospective occupation reported by respondents. The x-axis uses the average income score based on the five self-reported occupations of fathers closest to age 40. Both axes use income scores at the *occupation* × *race* × *South* level constructed using the 1970 Census.

Table C.7: Relationship between 5-year average of father's self-reported income and various other ways to measure father's income, 1997 PSID

	Income Score			Actual Income
	(1) Self-reported, in 1970	(2) Retrospective	(3) Self-reported, 1 year, age 40	(4) Self-reported, 1 year, age 40
Logged income	1.009	1.002	0.999	0.572
	[0.108]	[0.109]	[0.108]	[0.161]
Observations	898	898	898	898
R-squared	0.36	0.35	0.36	0.65

Notes: This table regresses the 5-year average of father's self-reported logged family income on four alternative ways to measure father's income level (denoted in the column headers). The sample used is the fathers of household heads ages 30–50 who provided a retrospective answer in 1997. We include fathers who can be located in an earlier wave of the survey and who had at least five years of available income and occupation information between the ages of 30–50. The dependent variable is the average of five years of father's logged family income closest to age 40. Column 1 uses the income score of the father's self-reported occupation around 1970 as the independent variable. Column 2 uses the income score corresponding to the retrospective answer provided by the household head about their father. Column 3 uses the income score corresponding to the father's self-reported occupation closest to age 40. Column 4 uses the family income of the father in one year closest to age 40. All income scores are constructed using the 1970 Census and vary at the *occupation* $\times$ *race* $\times$ *South* level. All estimates are weighted using 1997 cross-sectional weights.

Table C.8: IGE and Rank Coefficient Using Retrospective Adult Child's Answer and Self-Reported Father's Occupation, 1997 PSID

(a) Logged father's income

	(1) Retrospective	(2) Self-reported, 1 year	(3) Self-reported, 3 years
Father income score	0.564	0.564	0.585
	[0.096]	[0.085]	[0.088]
Observations	1,055	1,055	1,055

(b) Ranked father's income

	(1) Retrospective	(2) Self-reported, 1 year	(3) Self-reported, 3 years
Father income score	0.225	0.210	0.207
	[0.040]	[0.039]	[0.039]
Observations	1,055	1,055	1,055

Notes: This table reports estimates of the IGE and rank-rank coefficients from specifications that use respondents' retrospective answers about their fathers' occupations (column 1) and fathers' self-reported answers about their occupations in earlier survey waves (columns 2 and 3). The sample used is household heads ages 30–50 who provided a retrospective answer in 1997 and whose father can be located in an earlier wave of the survey. Column 1 uses the provided retrospective answers. Column 2 uses the self-reported occupation of fathers ages 30–50 roughly thirty years earlier (around 1970). The last column uses three years of self-reported occupations (between 1968–1972) and takes an average of the three corresponding income scores. All estimates use income scores at the *occupation* × *race* × *South* level constructed using the 1970 Census and are weighted using 1997 cross-sectional weights.

D Additional detail on data sources

D.1 Harmonizing surveys

We typically include a survey in the analysis if it meets two main conditions: First, it must ask survey respondents about their family income. And second, it must ask respondents about their fathers' occupation while they were growing up, and the available occupation codes must be able to be mapped to our coarsened occupations (discussed below). The surveys that meet these conditions usually also include other useful information, including demographic characteristics of the respondent (e.g., age, country of birth, education, occupation) as well as of the father and mother (e.g., education).

In the end, we have fifteen harmonized surveys:

- American National Election Studies (ANES),1956–1970
- Americans View Their Mental Health (AVTMH), 1957 & 1976
- General Social Survey (GSS), 1972-2018
- National Fertility Survey (NFS), 1970
- NLS Mature Women (NLSMW), 1968
- NLS Older Men (NLSOM), 1966
- NLS of Youth, 2002 [35]
- NLS Young Men (NLSYM), 1981 [35]
- NLS Young Women (NLSYW), 1988 [35]
- National Survey of Black Americans (NSBA), 1979–1980
- National Survey of Families and Households (NSFH), 1987–1988
- Occupational Changes in a Generation (OCG), 1962 & 1973
- Panel Study of Income Dynamics (PSID), 1997 & 2017 [36]

To minimize life-cycle bias, we restrict the sample to native-born respondents aged 30–50. We also include respondents in this age range for whom we do not know where they were born.[37] Finally, we restrict the sample to individuals whose race is recorded

[36] Note that the National Longitudinal Surveys can also be used as repeated cross-sections. For these three surveys, we select the cross-section to use by first observing the median age in the earliest cross-section of the survey. We then calculate the year in which the median age of respondents would be around 40. If the survey was not conducted in this year, we take the nearest survey year. Nevertheless, in these three NLS surveys (similarly to in the NLSMW and NLSOM) we typically use the first wave to collect demographic information about the respondent (e.g., sex, race, birth year, birthplace) as well as retrospective information about the parents' occupations and educational attainment. The one exception is that in the NLSYW survey, we collect information about the mother's occupation when the question was re-asked in 1978.

[36] Note that the PSID can be used as repeated cross-sections. We rely on the 1997 wave because this was the first year in which retrospective questions about parents' occupations were asked with sufficient detail and in which cross-sectional individual weights were available. We also then use the 2017 survey to bring in a new cohort of individuals ages 30–50. We intentionally exclude any 1997 respondents who appear again in 2017. Retrospective questions were only asked to household heads and their wives; whenever we have two respondents within a family with all of the available information for our analysis, we select a member at random.

[37] We exclude foreign-born respondents because we cannot know with certainty whether they grew up in or

as white or Black.[38]

Once we identify and clean these surveys, we pool them together for the analysis. An individual is in our baseline sample if he/she has an available family income, recorded race, region of birthplace/childhood (South vs. non-South), and father's occupation. Together, these four components allow us to measure the respondent's income level and compare it to his/her father's income score. The cohorts and samples in each survey are summarized in Figure D.1 and Table D.1.

D.2 Respondent family income

In all of our harmonized surveys, respondents are asked about their family income in that year. Some surveys provide the information in categories, while others provide exact numerical values. To be consistent in our coding, we rely on the bin structure of the surveys and assign respondents the midpoint of that category.[39]

For surveys that report exact income values, we replicate the bin structure for assigning respondents a family income value. In particular, we first find a survey that took place around the same time period and use that survey's bin structure as a template. We then assign individuals the midpoint of their corresponding bin.[40] Ultimately, we want to observe a roughly equal proportion of respondents in each bin. When the outlined procedure does not yield this result, we consider alternative bin structures (namely, the bin structure in other surveys) until we find a template that results in a relatively equal distribution.

Finally, for consistency, we ensure that each survey has roughly 10–12 bins for respondent family income. For surveys that have significantly more bins, we combine bins and assign respondents the midpoint of the new category (while simultaneously ensuring that each bin has roughly the same share of respondents).

D.3 Income scores for fathers

D.3.1 Coarsened occupations

We obtain father occupation from the respondent, who typically reports his/her father's occupation when the respondent was growing up or around 14–16 years old. In many

outside of the United States. Because we assign U.S.-based income scores to the father of each respondent and because the average income for the same occupation can differ across countries, we refrain from assigning income scores to the fathers of these respondents and thus exclude these father-children pairs.

[38] Respondents who are classified as Hispanic in surveys are re-classified as white unless there is additional information available on race. Respondents of other races, who comprise a tiny share of survey samples, are excluded from the analysis.

[39] The exception to this step is that for individuals who make the least (i.e., whose income falls in the bottom bin), we assign them 0.75 × the upper boundry of the category. For respondents who make the most (i.e., whose income falls in the top bin), we assign them 1.25 × the lower boundry of the category.

[40] For instance, because NSFH interviews took place in 1987 and 1988, we use the 1988 bins from the GSS as a template for the bin structure of family income for NSFH respondents.

of the surveys, we are also able to obtain analogous information for the mother's occupation.

Across all surveys, we harmonize occupations into 28 categories, corresponding to the main occupations in the American National Election Survey. The ANES occupation we use are:

- Accountants and auditors
- Clergymen
- Teachers
- Dentists
- Physicians and surgeons
- Engineers
- Lawyers and judges
- Social and welfare workers
- Nurses
- Other professional and technical occupations
- Semi-professional occupations
- Self-employed businessmen, managers, and officials
- Businessmen, managers, and officials
- Bookkeepers
- Stenographers, typists, and secretaries
- Other clerical workers
- Higher-status sales workers in "outside" sales
- Inside sales workers (e.g., salesmen, clerks)
- Lower-status sales workers in "outside" sales (e.g., peddlers, newsboys)
- Foremen
- Skilled craftsmen and kindred workers
- Semi-skilled operatives and kindred workers
- Protective service workers
- Private household workers
- Other service workers
- Farm laborers
- Non-farm laborers
- Farm operators

D.3.2 Constructing Census-based income scores

Because we want our baseline income scores to approximate the income of the fathers' generation, we restrict the decennial Census to individuals who resemble the survey respondents' fathers (Ruggles *et al.*, 2021). In particular, we restrict the sample to men who are between the ages of 30–50, whose race was recorded as either white or Black, and who had a child younger than 18 present in the household. For the 1950 Census, we also restrict the sample to men who were sample-line individuals (i.e., who were asked questions about income). We then build and use crosswalks that map the Census occupations into our 28 coarsened occupations.

Next, we calculate the median income in each occupation for individuals with certain characteristics. In our baseline specification, we calculate the median income by *occupation* $\times$ *race* $\times$ *South*, but we also vary this level of granularity in the robustness checks (e.g., by education level).[41] Our preferred measure of income is median household income, which sums the income of all family members within a household. In a robustness check, we also use median father's income, which only considers the income of fathers.

The 1940 income variable (i.e., wage and salary income) excludes income from self-employment and farming. We thus implement two main changes to our baseline income score following the approach in Collins and Wanamaker (2017). First, we use fathers ages 30–50 in the 1960 Census to calculate the ratio of farmer income to farm laborer income. We then use farm laborers' income in 1940 as well as these ratios to impute the 1940 income of farmers.[42] Second, we adjust the income of self-employed non-farm workers using a similar approach: we consider fathers ages 30–50 in the 1960 Census and compute ratios of median earnings for self-employed workers relative to wage-and-salary workers. We then impute the earnings of self-employed non-farm workers in 1940 using these ratios. Throughout the analysis, we use the same level of granularity to compute ratios as we do when constructing income scores. Our preferred ratios therefore vary at the *race* $\times$ *South* level.[43]

In the robustness checks, we calculate analogous income scores—at the *occupation* $\times$ *race* $\times$ *South* level—using the 1950, 1960, 1970, 1980, and 1990 Censuses. In all of these variations, we calculate the median of household income using the *inctot* (i.e., total personal income) variable.[44] Finally, we also calculate the median of the 1950 *occscore* variable—which reflects the median total income of all persons with that particular occupation in that Census—for the 28 coarsened occupations (with no additional variation at the race or region level). To ensure that these measures are comparable throughout the analysis, all income scores are reported in 1950 dollars.

Finally, we construct income scores for working mothers and for non-working fathers. For individuals who provided information about their mothers' occupations, we construct analogous income scores to our baseline income scores, but using mothers ages 30–50 in the 1940 Census who were heads of household. Moreover, certain survey respondents had a missing father occupation (i.e., their occupation was not mapped into one of our 28 categories) not because the respondent did not know what the oc-

---

[41] For education variations, we use five levels of education: less than 8th grade, 8th grade, some high school, completed high school, and at least some college.

[42] Throughout these calculations, we also follow Collins and Wanamaker (2017) and adjust farmer and farm laborer income measures upward to reflect the value of in-kind income.

[43] If there were fewer than 20 individuals in the 1960 Census cell (e.g., *occupation* $\times$ *race* $\times$ *South* $\times$ *education*), we rely on the median income of individuals in the broader group (*occupation* $\times$ *race* $\times$ *South*) to construct ratios.

[44] The 1950 Census only asked the income question to a small subset of the population. As such, for the 1950 *occupation* $\times$ *race* $\times$ *South* variation, if there were no respondents in the 1950 Census with a particular occupation, Southern residence, and recorded race, we impute the income value using the income of similar individuals (i.e., same race, same residence, similar occupation).

cupation was, but because the respondent reported that their father was not working (e.g., unemployed, retired). We assign an income to these non-working fathers using income ratios from the 1960 Census, similar to the adjustments for farmer and self-employed fathers described above. In particular, we calculate ratios at the *race* $\times$ *South* level in the 1960 Census using the median household income of working fathers and of non-working fathers (unemployed or out of the labor force). We then use these ratios and the median income score of working fathers in the surveys (in that *race* $\times$ *South* cell) to impute the income score for non-working fathers.

D.3.3 Alternative income scores

In the robustness checks of the paper, we consider a number of alternative income scores. The first variation we consider is one that uses alternative data sources (i.e., data not from the decennial Censuses). In particular, for non-farmer fathers, we use information on average earnings by occupation from the 1901 Cost of Living Survey (Preston and Haines 1991) and collapse this information to our coarsened occupations. We use the income of fathers ages 30–50 in the 1940 full-count Census to adjust these income values by race and Southern residence. For fathers who are farmers, we assign an income value using the 1900 Census of Agriculture. In particular, we use information on farm output and expenses from Merriam (1902) and follow the approach in Goldenweiser (1916) and Abramitzky *et al.* (2012) to calculate farmers' income by race and Southern residence. We additionally adjust these values by the share of farmers in that race and region that were owners (assuming that non-owners earn 50% of the estimated farm income). We refer to income scores that combine the 1901 Cost of Living Survey and the 1900 Census of Agriculture as our 1900-based income scores.

A second variation we consider is one in which we assign fathers an income score using the data sources that are closest in time to when the respondent is 10 years old. In particular, we assign cohorts born between 1910 and 1930 a weighted average of the 1900-based income scores and our 1940 baseline income scores, with the weights reflecting the number of years between when the respondent was 10 years old and 1940. For cohorts born between 1930 and 1940, we assign a weighted average of the 1940- and 1950-based income scores, with the weights once again reflecting the number of years between when the respondent was 10 years old and 1950. We continue this process for all respondents born in the 1940s–1970s cohorts using the income scores constructed using the 1950–1990 Censuses.

D.3.4 Assigning income scores to survey respondents' parents

As previously mentioned, we harmonize fathers' occupations (and mothers' occupation whenever available) into 28 coarsened categories. To do so, we construct crosswalks between the 1950 Census occupations and our coarsened occupations, as well as analogous crosswalks for the 1960, 1970, 1980, 2000, and 2010 Census occupations. If the occupations in a survey did not match the Census list of occupations, then we created survey-specific crosswalks between the available occupation codes and our coarsened occupations.

Once we finish coarsening occupations, we merge our income scores by father occupation, race, and whether the *respondent* grew up in the South. While our surveys provide father occupation, they do not report information on his race. We thus proxy father race with respondent race. Moreover, our surveys do not report the state or region in which the respondent's father worked when the respondent was growing up. We can, however, observe the region in which the respondent was born or grew up. We therefore use respondent residence in childhood/adolescence to proxy for father residence. Whenever we have information on both birthplace as well as childhood region, we use the latter to proxy for father residence.

Finally, whenever a father's occupation is unavailable but the occupation of the mother is provided, we merge in the corresponding income scores for mothers, again by mother's occupation, race, and whether the respondent grew up in the South.

D.4 Educational attainment

Our constructed measures of educational attainment always reflect years of schooling *completed*. In some surveys, respondent and father education are binned (i.e., "less than grade school," "grade school," "less than high school," etc.), while in other surveys they are categorical (i.e., 0-20+ years of schooling). To harmonize across surveys, we create two education variables.

The first binned variable assigns consecutive, ascending values as follows:

(0) no education (0 years)
(1) less than grade school (1-7 years)
(2) grade school (8 years)
(3) less than high school (9-11 years)
(4) high school (12 years)
(5) some college (13-15 years)
(6) college+ (16+ years)

In contrast, the second binned variable assigns *years of schooling* in the following manner:

(0) no education
(6) less than grade school
(8) grade school
(10) less than high school
(12) high school
(14) some college
(16) college+

We create these two variables for the respondent and for the respondent's father. Whenever available, we make similar variables for the respondents' mothers. Finally, we create indicator variables denoting high school and college completion for the respondent, for the father, and for the mother if possible.

D.5 Weighting scheme

We construct two types of weights for our analysis, a centered weight and a population-adjusted weight.

We begin by taking the provided weight in each survey and dividing it by its mean so that the weight has an average of 1 within a survey. For surveys that consist of repeated cross-sections (i.e., the ANES and GSS), we re-center the weight in each survey year. If a survey does not have a weight, we create a weight with all values set to 1. We then combine these re-centered weights into one variable: the centered weight.

The main weight we use in the analysis builds on the centered weight, but adjusts it further for population characteristics. In particular, because some of our surveys are not representative by race or sex, certain cohorts in the pooled dataset will not be nationally representative. We therefore adjust the centered weights so that the share of white men, white women, Black men, and Black women in each cohort (i.e., decade) is 44, 44, 6, and 6 percent, respectively.

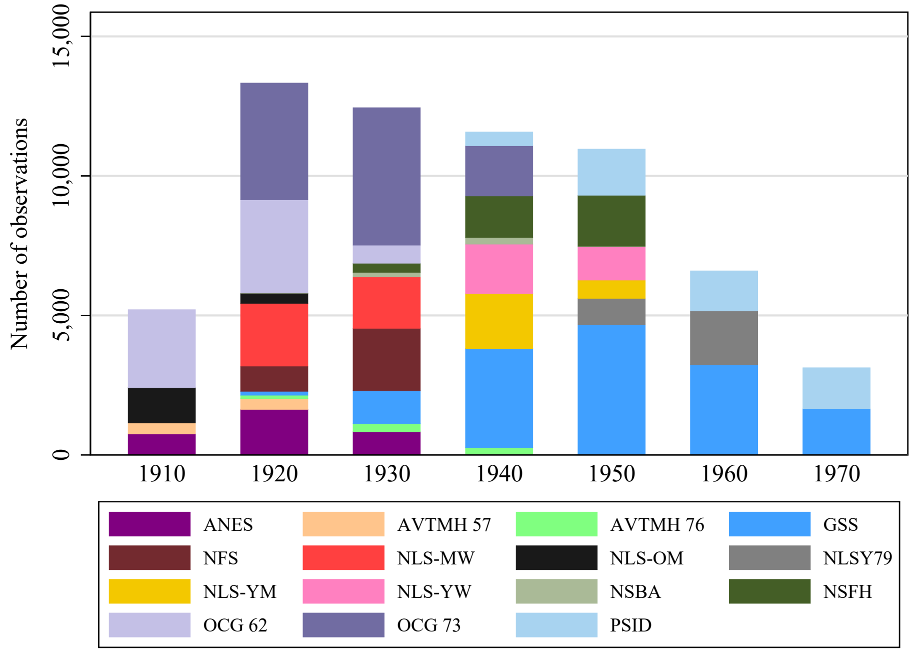
Throughout the analysis, we sometimes restrict the sample to certain respondents (e.g., individuals whose fathers are not farmers, individuals with available information on father's education). For these secondary samples, we also adjust the centered weight so that the share of white men, white women, Black men, and Black women in each cohort of that sub-sample is 44, 44, 6, and 6 percent, respectively.

D.6 Ranking respondents and their fathers

In addition to logging respondent family income and fathers' income scores, we also rank respondents and their fathers. In particular, we rank respondents relative to other survey respondents born in the same birth year. Similarly, we rank fathers relative to all other fathers with children born in the same year. Notably, we rank respondents and their fathers on the condition that we have a minimum of 100 observations in a given birth year for the relevant sample. Our baseline analysis sample ends up including individuals born in every year between 1911 and 1979. Finally, in our baseline approach, we use the population-adjusted weights when creating ranks.

Whenever we consider secondary samples of individuals, we re-rank respondents (and their fathers) so that individuals are compared to the other individuals in that sub-sample. We use the population-adjusted weights that correspond to that sub-sample when ranking.

Figure D.1: Survey data per birth cohort



Sources: This figure combines the 15 surveys, showing the number of respondents in each birth cohort in our baseline sample.

Table D.1: Cohorts and Sample in Each Survey

Survey	Cohorts	Sample
American National Election Survey, 1956–1970	1910–1930	Representative
Americans View Their Mental Health, 1957	1910–1920	Representative
Americans View Their Mental Health, 1976	1920–1940	Representative
General Social Surveys, 1977–2018	1920–1970	Representative
Occupational Changes in a Generation, 1962	1910–1930	Representative & male
Occupational Changes in a Generation, 1973	1920–1940	Representative & male
National Fertility Survey, 1970	1920–1930	Ever-married women ages 30–44
National Longitudinal Survey of Mature Women, 1967	1920–1930	Representative & female, ages 30–44
National Longitudinal Survey of Older Men, 1966	1910–1920	Representative & male, ages 45–50
National Longitudinal Survey of Young Women, 1988	1940–1950	Representative & female, ages 34–46
National Longitudinal Survey of Young Men, 1981	1940–1950	Representative & male, ages 30–40
National Longitudinal Survey of Youth, 2002	1950–1960	Representative, ages 37–45
National Survey of Black Americans 1979–1980	1920–1950	Representative & Black Americans
National Survey of Families and Households 1987–1988	1930–1950	Representative
Panel Study of Income Dynamics, 1997 & 2017	1940–1970	Representative

Notes: This table reports the cohorts and sample for each of the 15 surveys in our baseline sample. "Representative & male" and "Representative & female" refers to having representative samples by race within an all-male or all-female survey, respectively. "Representative & Black Americans" refers to representative samples (e.g., in terms of age groups) within the Black-American population.